# Cognitive load and working memory in multimedia learning:

# Conceptual and measurement issues

Øistein Anmarkrud[1], Anette Andresen[1], & Ivar Bråten[2]

[1]Department of Special Needs Education, University of Oslo, Norway.

[2]Department of Education, University of Oslo, Norway.

## Abstract

This article reviews contemporary research on multimedia learning that uses cognitive load theory as the major theoretical framework. In particular, we address the extent to which working memory has been conceptualized and measured in this research, what kind of subjective measures of cognitive load have been used and whether such measures are combined with other measures of cognitive load, and how results from subjective measures have been related to learning and achievement. The findings show that most of the reviewed studies did not include any clear conceptualization or measurement of working memory, used only general subjective measures containing one or very few items, and did not report findings consistent with the hypothesized relationship between cognitive load and multimedia learning. The findings are discussed in relation to the broader goal of improving research on cognitive load in the context of multimedia learning.


*Keywords:* Cognitive load, working memory, multimedia learning, subjective measures.

**Cognitive Load and Working Memory in Multimedia Learning:**

**Conceptual and Measurement Issues**

The 21st-century classroom is unprecedented in its use of multimedia, which involves combinations of various representations such as text (written or spoken), static graphics (e.g., pictures, illustrations, graphs, and diagrams), and dynamic graphics (e.g., animations and video; Butcher, 2014; Mayer, 2014). That is, new technologies, for example in the form of multimedia encyclopedias, simulation software, intelligent tutoring systems, or e-Courses, support the use of multiple formats that can be used to present information to students in flexible ways, with potential benefits for learning. Presenting instructional materials in the form of multimedia is often justified by referring to "the multimedia principle," basically saying that learning with words and pictures is more effective than learning with words or pictures alone (e.g., Butcher, 2014).[1]

Of note is that "multimedia" may be conceptualized more narrowly, such as simple combinations of text and graphic content within the same document, as exemplified by the work of Mayer and colleagues (e.g., Mayer, 1989; Mayer, Bove, Bryman, Mars, & Tapangco, 1996; Mayer & Gallini, 1990). However, multimedia also has come to be conceptualized more broadly, involving complex combinations of verbal (written or spoken) and pictorial (static or dynamic graphics) representations (Butcher, 2014; Mayer, 2014). Thus, in studies of multimedia learning, the complexity of multimedia may range from combinations of text and static graphics (e.g., Acarturk & Ozcelik, 2017; Chiu & Mok, 2017) to combinations of text,

[1] Importantly, the conditions of applicability of this basic multimedia principle are quite complex. Some of these conditions are discussed in a subsequent section termed *Design Principles of Multimedia Learning*.

audio files, and video files (e.g., Andrade, Huang, & Bohn, 2015; Chung, Cheon, & Lee, 2015) to combinations of texts, static graphics, video files, and several web pages (e.g., Cheng, Lu, & Yang, 2015). In particular, multiple representations distributed across multiple source documents may represent a complex combination in multimedia learning (List & Alexander, 2018). Multimedia learning may thus involve acquiring or constructing knowledge by processing combinations of verbal and pictorial representations at varying levels of complexity.

A substantial amount of educational research on multimedia learning uses cognitive load theory (Sweller, Ayres, & Kalyuga, 2011) to explain the potential benefits of using multimedia to enhance learning. At the core of cognitive load theory is the suggestion that multimedia learning can be best explained by examining how information is processed and stored, with particular attention being given to working memory and its role in learning (Schüler, Scheiter, & van Genuchten, 2011). Although research repeatedly has shown that multimedia may have several advantages for learning, quite a few studies fail to demonstrate positive outcomes (e.g., Dillon & Jobst, 2005; Lunts, 2002; Scheiter & Gerjets, 2007). In their extensive review, Scheiter and Gerjets (2007) identified several reasons for the ambiguity of results, arguing that the same features of digital environments that can promote learning may also be detrimental. For example, information represented in different ways has to be integrated by the learner. Previous work on the integration of verbal and visual information when reading traditional texts has shown that this integration can be challenging (Chan & Unsworth, 2011), and there are indications that meaning-making in multimedia reading environments poses even greater processing demands on the reader (Chan & Unsworth, 2011; Hartman, Morsink, & Zheng, 2010; Kamil & Chou, 2009). In brief, multimedia learning can represent unique opportunities for learning but can also challenge the learner’s working-memory capacity (Kalyuga, 2007; Scheiter & Gerjets, 2007).

The purpose of this article is to review contemporary research on multimedia learning that uses cognitive load theory as the major theoretical framework, with an emphasis on the conceptualization and measurement of working memory and the use of subjective self-report measures of cognitive load in this research. Specifically, we wanted to examine the extent to which working memory is clearly conceptualized and measured, what kind of subjective measures are used and whether such measures are combined with other measures of cognitive load, and how results from subjective measures of cognitive load are related to learning and achievement.

The reason we focus on the conceptualization and measurement of working memory is that working memory is considered a theoretical cornerstone in this area of research (Schüler et al., 2011; Sweller et al., 2011), and that careful attention to this variable presumably will facilitate the interpretation of results (e.g., by controlling for potential differences in working memory capacity between intervention and control groups; de Jong, 2010). Also, we specifically focus on the use of subjective measures of cognitive load because the psychometric properties of such measures may be open to question, at least when consisting of very few items and not combined with other types of cognitive load measures (de Jong, 2010; Naismith, Cheung, Ringsted, & Cavalcanti, 2015). In this regard, reliable and valid measures of cognitive load may be needed to better understand the cognitive mechanisms underlying effects (or lack of effects) of trying to reduce cognitive load.

In summary, our objective is to contribute to an improvement of research on cognitive load in the context of multimedia learning. That is, identifying potential issues with the conceptualization and measurement of working memory and the use of subjective measures of cognitive load in the latest multimedia research framed by cognitive load theory may provide an impetus for researchers in this area to refine their research designs and measurements by taking such lingering issues into consideration. In the following, we first discuss the concept of working memory and, in particular, how working memory is represented within cognitive load theory. Next, we describe and discuss approaches to measuring cognitive load in multimedia learning. Based on this background analysis, we further specify the research questions guiding our review, present our results, and discuss their implications for research on cognitive load in the context of multimedia learning.

## Working Memory and Cognitive Load

### Working Memory

Working memory, often conceived as a “mental workspace”, can be defined as a processing resource with limited capacity involved in the storage of information while simultaneously manipulating information for brief periods of time (Alloway, 2009; Baddeley & Logie, 1999; Swanson & Alloway, 2012). There is an extensive research literature documenting a significant relationship between working memory capacity and academic achievement in domains such as reading (e.g., Cantin, Gnaedinger, Gallaway, Hesson-McInnis, & Hund, 2016; Sesma, Mahone, Levine, Eason, & Cutting, 2009; Swanson & Jerman, 2007), mathematics (e.g., Gathercole, Pickering, Knight, & Stegmann, 2004; Purpura & Ganley, 2014; Swanson, Jerman, & Zheng, 2008), and multimedia learning (e.g., Austin, 2009; Pazzaglia, Toso, & Cacciamani, 2007; Skuballa, Schwonke, & Renkl, 2012). Although several models of working memory have been proposed in the literature (for a review, see Miyake & Shah, 1999), Baddeley’s (e.g., Baddeley, 1992, 1999; Baddeley & Logie, 1999) model of working memory is widely used in research linking working memory to academic performance (Swanson & Alloway, 2012). According to this model, working memory consists of four components: the central executive, the phonological loop, the visuo-spatial sketchpad, and the episodic buffer (Baddeley, 2000).

The central executive functions as an attentional control system that regulates the three other subsystems. It has several tasks in Baddeley’s model: (a) it monitors and directs processing in the other subsystems and links the results of this processing to long-term memory, (b) it distributes attention between tasks and assigns attention to incoming stimuli, (c) it transfers information to subsystems, (d) it updates and regulates the content of working memory, and (e) it codes representations for time and place. The phonological loop is the subsystem processing auditory information, including processing of both spoken and written words. The visuo-spatial sketchpad contains two components: the visual component addresses the characteristics of objects such as size, color, and shape, while the spatial component handles relational or spatial information and the control of movements. Finally, the episodic buffer is considered a limited-capacity storage system that temporarily stores and integrates information from the phonological loop and the visuo-spatial sketchpad with each other and with prior knowledge. Of note is that the idea of separate processing systems for verbal and visual information is not unique to Baddeley's model. For example, Paivio's (1971, 1986) classic dual coding theory provides a detailed explanation of how verbal and visual information is processed, integrated, stored, and recalled in the human mind.

The multimodal storage system described by Baddeley can be assumed to play a pivotal role in processing multimedia content. More specifically, three implications of Baddeley's model of working memory can be derived with respect to learning from multimedia (Schüler et al., 2011): (a) a reduced or limited working-memory capacity should have negative effects on multimedia learning, (b) these negative effects should hinge on which of the subsystems is affected by the resource limitations, and (c) limitations in the central executive and episodic buffer should have a negative effect on the integration of verbal and pictorial information. Given that working memory is an essential concept in cognitive load theory and research on multimedia learning, it is relevant to question the extent to which influential models of working memory, such as Baddeley's (e.g., 1999, 2000) model, are represented within cognitive load theory.

**Cognitive Load Theory**

Cognitive load theory was developed to examine the limitations of cognitive processing in relation to learning (Chandler & Sweller, 1991; Paas, Renkl, & Sweller, 2004; Sweller et al., 2011). At the core of the theory is an information-processing model of human cognitive architecture in which novel information must be processed by working memory before it can be stored in long-term memory. According to the central principle of cognitive load theory (Sweller, 2010; Sweller et al., 2011; van Merriënboer & Sweller, 2005), working memory capacity is limited, and instructional design should therefore aim to reduce unnecessary working memory load to free capacity for learning-related processing. Otherwise, exceeding working memory capacity could result in inhibited learning due to cognitive overload (Kalyuga, 2011; Leppink, Paas, van der Vleuten, van Gog, & van Merriënboer, 2013).

More specifically, cognitive load theory postulates that any learning task imposes three different categories of cognitive load on working memory: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load (Paas, Renkl, & Sweller, 2003; Paas et al., 2004; Sweller et al., 2011; van Merriënboer & Sweller, 2005). Intrinsic cognitive load refers to the inherent difficulty and complexity of any learning task. Some tasks (e.g., multiplication with three-digit numbers) have a higher inherit difficulty than other tasks (e.g., the calculation of "3 - 1"), regardless of the instructional approach. Hence, complex learning tasks impose a higher cognitive load on working memory than less demanding tasks. However, intrinsic cognitive load also depends on the prior knowledge of the learner, hence it is not solely a feature of the instructional content. Extraneous cognitive load, on the other hand, does not concern the complexity of a task but how the learning material is presented or the type of instructional activities undertaken. Extraneous cognitive load is thus strongly influenced by the instructional approach, for example the completeness of explanations or the integration of instructional materials. Finally, germane cognitive load refers to the load placed on working memory by learning, as when relating information from long-term memory to new information.

Of note is that the term “germane cognitive load” has been questioned lately. Thus, in their recent update of cognitive load theory, Sweller and colleagues (2011) preferred the term “germane resources” because unlike intrinsic and extraneous cognitive load, germane load is not imposed by the nature or structure of the learning material. Germane cognitive load instead arises from the process in which the learner relates relevant information from long-term memory to information from the present learning task (Sweller, 2010; Sweller et al., 2011). Hence, it pertains to the working memory resources allocated to dealing with intrinsic cognitive load (Kalyuga, 2011; Sweller, 2010).

## Design Principles in Multimedia Learning

Empirical studies have suggested that working memory load in learning with multimedia could be reduced by instructions that follow the principles of cognitive load theory (Sweller et al., 2011). Although people can be assumed to learn more from words and pictures than from words alone, not all multimedia presentations are considered equally effective (Mayer, 2005, 2009). According to Mayer (2005, 2009), instructional material should build on the knowledge of how the mind works and how people process information. Thus, evidence-based theory should guide the construction of multimedia instructional messages.

Important principles are that meaningful learning from multimedia is more likely to occur when corresponding words and pictures are presented near each other instead of far from each other on the page or screen (the spatial contiguity principle) and that the learning material should include cues to highlight how it is organized (the signaling principle; Mayer, 2005, 2009; Mayer & Fiorella, 2014). For example, splitting learners’ attention between two or more sources that have to be integrated is likely to have a negative effect on learning due to higher cognitive load compared to an environment where the information sources are spatially integrated (Schnotz & Kurschner, 2007). In addition, the redundancy principle posits that including additional information sources superfluous or unnecessary for understanding is likely to be detrimental to performance (Mayer, Heiser, & Lonn, 2001). Of note is that this effect may be moderated by learners' prior knowledge and expertise. For example, Kalyuga, Chandler, and Sweller (1998) found that high-expertise learners using a diagram did not need text-based information to be physically integrated with the diagram; in fact, they also experienced reduced cognitive load when the text was eliminated. Thus, within multimedia learning, the redundancy principle states that learning material should not combine text and pictures for learners for whom one of the representations is sufficient to build a mental representation because the redundant information requires mental effort without enhancing learning (Mayer, 2005). That experts may suffer from what novices find beneficial in this regard is called the expertise reversal effect. Of note is that this effect is less likely to appear at the extreme ends of the complexity spectrum, that is, when learning tasks represent very low or very high intrinsic cognitive load (Sweller et al., 2011).

## Working Memory within Cognitive Load Theory

Although cognitive load theory often refers to Baddeley's (1992, 1999; Baddeley & Logie, 1999) model of working memory as an important foundation, it has also been argued that the link between cognitive load theory and models of working memory is "… loose and appears at least partly inconsistent" (Schüler et al., 2011, p. 392). In particular, this inconsistency concerns the central executive that is the core of Baddeley's model. Cognitive load theory has considered the existence of two modes of processing (i.e., auditory and visual) in working memory since the mid-1990s (e.g., Mousavi. Low, & Sweller, 1995; Tindall-Ford, Chandler, & Sweller, 1997). Reflecting this distinction, in the latest extensive presentation of cognitive load theory, Sweller et al. (2011) described working memory as consisting of partly independent processors "… that correspond to the modality of information" (p. 44). However, Sweller et al. (2011) argued that knowledge stored in long-term memory acts as a de facto central executive. If the notion of a knowledge-independent central executive were right, what would then determine its actions, asked Sweller et al. (2011). In their view, the answer to this question would be a hierarchy of central executives where the second controls the first, the third controls the second, and so forth. Hence, the idea of a knowledge-independent central-executive is not considered logically viable. As an alternative, Sweller et al. (2011) suggested that only the combination of knowledge stored in long term memory and a "… random generate and test procedure is required for a functioning natural information system" (p. 35). It is therefore reasonable to conclude that although Sweller et al.'s (2011) perspective on working memory is in line with Baddeley's model as far as modality specific processing components are concerned, there are substantial differences regarding how these subcomponents are regulated and governed.

## Measurement of Cognitive Load

In the early days of cognitive load theory, cognitive load was predominantly measured using indirect measures such as time on task or error rates (Sweller et al., 2011). Although these approaches have been replaced by more direct measures to a large degree, they are still occasionally used to assess cognitive load in multimedia contexts (e.g., Mahdjoubi & A-Rahman, 2012). Three main categories of direct measures of cognitive load have been established in the literature: dual-task methodology, physiological measures, and subjective ratings (Leppink et al., 2013; Naismith & Cavalcanti, 2015; Paas, Tuovinen, Tabbers, & van Gerven, 2003; Sweller et al., 2011).

### Dual-Task Methodology

Dual-task methodology includes a secondary task in combination with a primary task. For example, to examine the load imposed by a primary task (e.g., understanding UV radiation by studying a website containing multiple representations), participants are asked to press the space key on the computer keyboard every time a specific sound is played (secondary task). If the primary task imposes a high cognitive load, performance on the secondary task (e.g., response time) is expected to deteriorate if the primary and secondary tasks tap the same working memory resources (Brünken, Plass, & Leutner, 2003, 2004; DeLeeuw & Mayer, 2008). In the majority of research on cognitive load using dual-task methodology, the secondary task is dissimilar and requires less working memory than does the primary task (Sweller et al., 2011). Examples include pressing a pedal when a particular tone sounds (Brünken et al., 2004; Marcus, Cooper, & Sweller, 1996), pressing a key on a computer keyboard when a letter changes color (Brünken, Steinbacher, Plass, & Leutner, 2002), or noticing when a button is illuminated (van Gerven, Paas, van Merriënboer, & Schmidt, 2006). Sweller (1988) described and examined another, more complex, secondary task. He proposed that using problem solving as a strategy for learning involves two separable tasks: solving the problem and learning from problem solving. Sweller (1988) hypothesized that if learners treat problem solving as the primary task, learning becomes a secondary task, and poorer learning could result from high-demand problem solving.

**Physiological Measures**

The development of more sophisticated technologies has led to increased interest in physiological measures of cognitive load, and evidence has started to emerge that techniques such as cognitive papillary response, functional magnetic resonance imaging (fMRI), and electroencephalography (EEG) have considerable merit (Paas, Ayres, & Pachman, 2008; Sweller et al., 2011). Within the context of multimedia learning, there is also an increased use of eye-tracking methodology to examine changes in cognitive processing, and it has been argued that an increased length of eye fixation and greater pupil size are related to increased cognitive load (Chuang & Liu, 2012; Underwood, Jebbet, & Roberts, 2004; van Gog & Jarodzka, 2013). Yet, eye-tracking data are not self-explanatory; there can be several reasons why a participant was looking somewhere for a certain amount of time or in a certain order (e.g., task instructions). Eye-tracking data should consequently be used in combination with other measures of cognitive load (e.g., Kok & Jarodzka, 2017; van Gog & Jarodzka, 2013).

**Subjective Measures**

The use of subjective measures is by far the most common approach to measuring cognitive load in the literature. In their recent systematic review of the validity of cognitive-load measures in simulation-based training, Naismith et al. (2015) found that 71% (34 of 48) of the studies reviewed used subjective measures of cognitive load, corroborating findings from previous reviews of the literature (e.g., Paas, Tuovinen, et al., 2003). The assumption underlying the use of subjective measures is that people are able to evaluate their own cognitive processes and provide a valid estimate of the cognitive load imposed by a task. This evaluation typically involves asking participants to rate their cognitive load on one or more items with Likert-type scales, with the Paas scale (1992) of mental effort being the most used and cited subjective measure (Naismith & Cavalcanti, 2015). The original Paas (1992) scale consists of one single item where participants rate their mental effort on a nine-point scale ranging from *very, very low mental effort* to *very, very high mental effort*. Although Paas and colleagues reported good reliability and sensitivity for this measure when it was introduced (Paas, 1992; Paas & van Merriënboer, 1994; Paas, van Merriënboer, & Adam, 1994), more recent studies of cognitive load using subjective measures have been criticized for not reporting any reliability or sensitivity data (Naismith & Cavalcanti, 2015; Naismith et al., 2015; Paas, Tuovinen, et al., 2003). Moving beyond such reliance on a single item measure, recent attempts to measure different types of cognitive load (i.e., intrinsic, extraneous, and germane load) with multiple-item scales have showed promising results (e.g., Leppink et al., 2013; Sewell, Boscardin, Young, Cate, & O'Sullivan, 2016).

Another issue related to the use of subjective measures is how cognitive load has been conceptualized and operationalized in empirical research. Both perceived difficulty and expended effort have been widely used as proxies for cognitive load in various subjective measures without any clarification of the difference between the two constructs. For example, de Jong (2010) found in a review that perceived difficulty and mental effort were used interchangeably in several studies, with perceived difficulty being the term used in measures of cognitive load, and mental effort being the term used when presenting and discussing the results. In addition, de Jong (2010) found instances in which both terms were used separately in questionnaires but then combined into one construct in the analyses without further clarification. The relationship between the two constructs has been empirically examined to a minor degree. However, there are some studies indicating that perceived difficulty and mental effort are different constructs with distinct consequences for the measurement of cognitive load (Ayres & Youssef, 2008; Schmeck, Opfermann, van Gog, Paas, & Leutner, 2015). For example, although measures of perceived difficulty and effort are often correlated, the learner may be unable or unwilling to put in substantial effort if a task is too demanding (Sweller et al., 2011). Cognitive load has also been operationalized through terms such as pressure during task (Cheng et al., 2015), mental load (Chen & Wu, 2015), frustration (Huang et al., 2015) and mental demand (Kizilcec et al., 2015). In other words, there is a lack of consistency in how cognitive load has been operationalized in subjective measures, which makes it difficult to interpret empirical findings.

Issues of timing also are evident in that subjective measures of cognitive load have been administered at different points of time during problem solving (Schmeck et al., 2015; van Gog, Kirschner, Kester, & Paas, 2012). Some studies (e.g., Dindar, Yurdakul, & Dönmez, 2015; Paas & van Merriënboer, 1994; Tabbers, Martens, & van Merriënboer, 2004) have applied one item examining the perceived difficulty or expended effort repeatedly after every task in a learning situation, with this approach often referred to as immediate ratings of cognitive load. In other studies (e.g., Inan et al., 2015; Kalyuga, Chandler, Touvinen, & Sweller, 2001; Leutner, Leopold, & Sumfleth, 2009), one or more items have been administered after the end of the full learning session in an approach referred to as delayed ratings of cognitive load. Studies that have compared immediate and delayed ratings of cognitive load have consistently found that delayed ratings yield ratings higher than the average of repeated immediate ratings made during problem solving (Schmeck et al., 2015; van Gog et al., 2012).

A single delayed item at the end of a learning session is easy to administer, but it can be challenging to determine what such a rating really measures. Some participants might try to estimate an average of their cognitive load across tasks, while others judge their cognitive load based on the last task, and still others base their judgment of cognitive load on the most demanding or complex task they worked on. Although all of these possibilities (and different combinations of them) are plausible, there is little empirical evidence available. However, Schmeck et al. (2015) found that delayed ratings seemed to be more influenced by the most challenging tasks in a learning situation, rather than by when these challenging tasks are administered (e.g., at the beginning or the end of a learning session).

**The Potential Intrusiveness of Subjective Measures**

To be able to give accurate delayed judgments of cognitive load after a series of tasks, learners would have to allocate working memory capacity to monitoring cognitive load for the various tasks while trying to solve them, store information from this monitoring in long-term memory, and retrieve this information from long-term memory back into immediate memory to answer the question(s) about cognitive load. On the other hand, the use of repeated measures of cognitive load after every task requires participants to consider the cognitive load of the task they just finished, drawing on information that is probably still in working memory (Schmeck et al., 2015; van Gog et al., 2012). A question that remains open, however, is the potential intrusiveness of repeated measures (Schmeck et al., 2015). If participants are told they will be asked questions about perceived difficulty or expended effort after every task, the measurement of cognitive load could be considered a secondary task. This means that explicitly encouraging participants to allocate working memory capacity to monitor their cognitive load during problem solving could lead to at least two outcomes when working on complex tasks: (a) little available working memory for the accurate judgment of cognitive load, resulting in inaccurate judgments of task difficulty or expended effort; or (b) poorer learning because the combined load from the task and the monitoring of cognitive load exceeds available working memory capacity.

Moreover, it can be assumed that the risk of intrusiveness of repeated measures of cognitive load would be particularly pronounced for participants with low working memory capacity because the working memory resources of these participants would be more easily depleted by the combined requirements of task completion and cognitive load monitoring (cf., Chen, Castro-Alonso, & Paas, 2018). As a consequence, it seems essential that studies using subjective measures of cognitive load measure and control for participants' working-memory capacity, as well as describe the procedures and instructions used for the subjective measures of cognitive load very carefully to make it clear whether there is any extraneous load imposed by the measure of cognitive load itself. Of note is that the potential intrusiveness of the cognitive load measures discussed above may be less dependent on the number of items than on the secondary task introduced by such measures. Thus, although only one single question about mental effort may be asked, cognitive load measurement could intrude upon judgment of cognitive load and/or learning because it may require that learners allocate working memory resources to monitor their cognitive load during learning task completion.

**The Present Review**

On this backdrop, we set out to review recent research using subjective measures of cognitive load when examining learning in a multimedia context. For several reasons, we restricted our review to studies published between 2014 and 2017. First, although subjective measures of cognitive load are widely used, there has been a discussion in the literature about whether such measures are sufficiently sensitive to discriminate between the different types of cognitive load and whether the subjective perceptions of task difficulty and mental effort are related to the actual cognitive load (e.g., Ayres, 2006; Brünken et al., 2003; van Gog & Paas, 2008). We therefore sought to determine to what extent the most recent published research on multimedia learning using subjective measures of cognitive load had responded to this critique by using scales with greater specificity and/or more items and by evaluating how cognitive load measured by means of subjective rating was related to learning.

Second, although the construct of working memory is arguably a cornerstone of cognitive load theory, the correspondence between influential models of working memory, such as Baddeley's (2000, 2001) model, and cognitive load theory is not straight forward (Schüler et al., 2011). In particular, Sweller et al.'s (2011) extensive and updated presentation of cognitive load theory presented a detailed and thorough explanation of the concept of working memory from a cognitive-load perspective. Therefore, we wanted to investigate whether the centrality of working memory within cognitive load theory was reflected in the most recent research on multimedia learning that used subjective measures of cognitive load.

Finally, in the last decade, cognitive load theory has been the subject of strong criticism regarding methodological and measurement issues (de Jong, 2010; Gerjets, Scheiter, & Cierniak, 2009) and conceptual clarity (de Jong, 2010; Schnotz & Kürschner, 2007). In particular, de Jong (2010), based on a thorough review of the most cited cognitive load research, highlighted the need to include multi-item subjective measures of cognitive load, preferably in combination with other types of measures of cognitive load. Moreover, de Jong (2010) highlighted the need to include measures of working memory as a control variable because "it is not enough to know the demand (which could mean the effort applied); the system's capacity must also be known in order to determine (over)load" (p. 123). Hence, we wanted to examine whether such critique and recommendations had been answered with more rigorous and comprehensive measurements as well as more conceptual clarity in the most recent research on cognitive load in a multimedia context.

Specifically, our review was guided by the following three questions.

1. How is working memory conceptualized in contemporary cognitive-load research on multimedia learning, and to what degree is working memory assessed in this research?

2. How are subjective measures of cognitive load used in this research, and are subjective measures combined with other measures of cognitive load?

3. How are the results from subjective measures of cognitive load related to learning and achievement?

## Method

### Search Strategy

We worked with a research librarian at our institution to develop a search protocol. The search was conducted in October 2017 and included four databases that represented both educational and psychological domains: ERIC, Social Sciences Citation Indexes, PsycINFO, and Web of Science. Search terms for cognitive load included cognitive load, mental load, workload, and mental effort. These search terms were combined with "multimedia" and/or "multimedia learning". We also performed manual searches of highly relevant journals for articles about cognitive load and multimedia learning. These journals included *Learning and*

*Instruction, Computers in Human Behavior*, *Journal of Science Education and Technology*, *Instructional Science, Applied Cognitive Psychology, Computers and Education, Journal of Computer Assisted Learning, Journal of Educational Psychology,* and *Journal of Experimental Psychology: Applied*. Both the search in the databases and the manual searches were restricted to articles published between January 1st, 2014 and October 1st, 2017.

**Study Selection and Data Extraction**

We used a multistage approach to select relevant studies for the review. Whereas no particular study design or outcome measure was required, studies had to meet the following inclusion and exclusion criteria:

1. Citing of cognitive load theory: The studies had to explicitly cite and build upon cognitive load theory as a major theoretical framework.

2. Measure of cognitive load: A subjective measure of cognitive load would have to be used in the study, either as the only measure of cognitive load or in combination with other types of measures (e.g., secondary task or physiological measures of cognitive load).

3. The study would have to address learning in a formal educational context by means of multimedia (e.g., by using tablets, simulation software, or educational computer games).

4. Population: Participants should have no identified cognitive impairment or learning disability.

Next, we developed a coding scheme to extract the following information from the selected studies:

1. Study aim, including main research question(s).

2. Method, including description of participants (age, educational background, and sample size), study design, multimedia tasks, measure(s) of cognitive load, procedures for measuring cognitive load, and measure of working memory (if measured).

3. Conceptualization of working memory in literature review, including definition of working memory and references for this definition.

4. Findings regarding the relationship between cognitive load and learning, categorized as negative (i.e., high cognitive load was associated with low learning), positive, or neutral.

5. Findings regarding the relationship between cognitive load and working-memory capacity, categorized as negative (i.e., high working memory capacity was associated with low cognitive load), positive, or neutral.

## Results

[Figure 1 about here]

Using the search terms described above, the initial search in the four databases yielded 381 results (see Figure 1). After the removal of 65 duplicates, the abstracts of the remaining 316 articles were examined by the first and second authors in light of the four inclusion and exclusion criteria, which resulted in 63 articles being removed from the review because they were theoretical papers or examined cognitive load in populations with learning disabilities. Based on a preliminary full-text examination of the remaining 253 articles, also conducted by the first and second authors in collaboration, 189 more articles were removed because they did not meet the inclusion and exclusion criteria. Specifically, 103 articles did not use subjective measures of cognitive load, 64 articles did not use cognitive load theory as an explicit theoretical framework, and 22 articles did not examine multimedial learning in a formal educational context. Manual searches of the journals mentioned above did not result in additional studies being identified. The final set of 64 articles, including 73 studies, were coded by the first author using the previously described coding system. Table 1 provides detailed descriptions of the 73 studies included in the review, whereas Table 2 summarizes the results across studies by year of publication. These studies represented all educational levels, with 44 studies including college or university students, 19 studies including high-school students, and 10 studies including elementary or middle school students.

[Tables 1 and 2 about here]

**Working Memory**

As can be seen in Table 2, only 20 (31.25%) of the articles included any conceptualization or clear definition of working memory by referring to classic sources in the working memory literature, such as Baddeley and colleagues (e.g., Baddeley, 1986; Baddeley, 2001; Baddeley & Hitch, 1974), or to Paivio's (e.g., 1971, 1986) work. Interestingly, there were no references to a cognitive-load perspective on working memory, such as that of Sweller et al. (2011). In the remaining 44 articles, working memory was often more or less explicitly described as limited in capacity through descriptions such as "… cognitive load exceeds working memory" (Cheng et al., 2015, p. 129), "… limited working memory resources" (Hsu, Gao, Liu, & Sweller, 2015, p. 114), or "… humans' limited working memory" (Huang, Chen, Wu, & Chen, 2015, p. 159), without further clarification, explanation, or reference to theories of working memory. Only 10 (13.70%) of the 73 studies included a measure of working memory in the research design. In most instances, this working memory measure was used to control for systematic differences between experimental and control groups, and not to examine the relationship between working memory and cognitive load.

**Measurement of Cognitive Load**

Based on the inclusion criteria, all studies used some form of subjective self-report to measure cognitive load. The majority of studies (58, 79.45%) used delayed ratings; that is, participants were given one or more items targeting cognitive load after a learning or problem solving phase. The remaining 15 studies used immediate ratings of cognitive load; that is, participants were given one or more items during a learning or problem solving phase.

There was also some variation concerning the number of items used to measure cognitive load. As many as 31 studies (42.47%) used one single item. However, in 16 of these studies, this item was used repeatedly during the learning phase. Further, 21 (28.77%) of the studies used two items, 13 studies (17.81%) used three or four items, and seven studies (9.59%) used five or more items when measuring cognitive load. A total of 35 studies (47.95%) used the original Paas (1992) one-item scale, either alone or in combination with other measures.

Although the measures used in these studies were similar in that they consisted of one or more items that participants rated on a scale, there was large variation in how the concept of cognitive load was operationalized in the measures. The majority of the studies included one or more items asking participants to rate their “mental effort” (53, 72.60%) and/or “perceived difficulty of the task” (36, 49.32%). Of note is that some studies (e.g., Hsu et al., 2015) referred to the original Paas (1992) scale, in which “mental effort” is used in the item, but still used “cognitive load” in their own measures. As summarized in Table 2, participants were to a lesser degree asked to rate their “cognitive load”, “mental load”, “pressure during problem solving”, “frustration”, “mental demand”, “perceived complexity”, “clarity of learning material”, “perceived learning”, and “expended attention”.

Despite explicit critique in recent years towards the use of subjective measures of cognitive load without reporting any reliability or sensitivity data from own sample, only 15 of the 73 studies we reviewed (i.e., 20.55%) provided such information. In many instances, authors did refer to the results of reliability analyses conducted by the developers of particular measures, without discussing the validity of those analyses when measures were translated and used in other cultural contexts. Concerning the instructions given to participants before they rated their cognitive load, the information provided in the 73 studies was not clear. That is, in the majority of the studies, it was not possible to determine from the description of procedures and instruments whether participants were told that they would receive questions about their cognitive load during or after the learning or problem solving phase. Only four of the studies (5.48%) provided information that made it possible to infer that participants actually knew that they should pay attention to their cognitive load during learning or problem solving. For example, although it was not explicitly stated, it was possible to infer that the participants in Dindar et al.'s (2015) study actually received such information, as the participants practiced "… how to respond to the Likert scale for reporting their CLs…" (p. 155).

Only 10 of the studies (13.70%) combined subjective rating of cognitive load with another method of measurement. The results concerning the relationship between the subjective measure of cognitive load and the additional measure were ambigious. Thus, five of the studies (e.g., Dindar et al., 2015; Korbach et al., 2017) did not find any statistically significant relationship between scores on their subjective measure and an additional measure. For example, Korbach et al. (2017) combined their subjective measure of cognitive load with a secondary task where participants tapped a predefined rhythm with their feet. The results showed that the accuracy of the tapping was related to learning, but this was not the case for the scores on the subjective measure of cognitive load. Further, there was no statistically significant relationship between scores on the subjective measure and the secondary task. In one of the studies (Lee, 2014), the results were less clear. In that study, a subjective measure was combined with electroencephalography (EEG), and a statistically significant correlation was found between perceived difficulty of task (subjective measure) and beta waves in the T3 area, while no such relationship was found between mental effort (subjective measure) and beta waves. In the four remaining studies, statistically significant relations between subjective and additional measures were found. For example, Marefat et al. (2016) combined their subjective measure of cognitive load with a secondary task where participants pressed the space key on a computer keyboard whenever a large letter at the bottom of the computer screen changed color, finding that both measures of cognitive load predicted learning.

**Relationship between Cognitive Load and Other Measured Variables**

As described previously, a central idea in research of multimedia learning from a cognitive-load perspective is that a good instructional design should reduce cognitive load, and learning should consequently improve because more working memory capacity is available. Among the 73 studies, 32 (43.84%) presented results consistent with this expectation. As an example, Yung and Paas (2015) examined the effect of adding visual representations to mathematical problems in a computer-based multimedia instructional context developed to teach students arithmetic operations. A control group used the same multimedia materials, but in the control condition, there were no visual representations. The authors hypothesized that supporting students' emerging arithmetic understanding with visual representations would lead to reduced cognitive load and better processing of the learning materials with consequent higher learning performance. The results were consistent with their hypothesis: participants in the condition with visual representations rated their cognitive load as lower during the learning phase and obtained better results on the posttest compared to participants in the control condition.

In other studies, however, the intervention in question resulted in better learning but there was no statistically significant effect on cognitive load. As an example, Inan et al. (2015) investigated the effects of modality on learning from multimedia instruction. In that study, participants used an interactive computer-presented diagram to learn about places of articulation in human speech. Two versions of the computer diagram were developed: one version in which the diagram (cross-section of the human speech system) included text descriptions explaining each point in the interactive diagram and one version using audio descriptions. The results showed that condition had a statistically significant effect on learning (participants in the text condition outperformed those who studied with audio descriptions), but there was no statistically significant effect on cognitive load.

In yet other studies, the experimental condition did not result in better learning but did have an effect on cognitive load. Of note, in Dindar et al. (2015), there was no effect of condition on the subjective ratings of cognitive load, but there was a statistically significant effect of condition on cognitive load measured using a secondary task. Additionally, there was no statistically significant correlation between the two measures of cognitive load.

# Discussion

In this review, we examined a sample of recent studies of multimedia learning framed by cognitive load theory that measured cognitive load by means of subjective ratings. In the following, we lay out and discuss the key findings and relate them to the broader goal of improving research on cognitive load in the context of multimedia learning.

## Working Memory in Recent Cognitive Load Research

The vast majority of the studies in this review did not refer to any conceptualization of working memory or provide any explicit definition of this construct. In one sense, this is somewhat surprising because working memory and cognitive load theory are by nature highly related (de Jong, 2010). Indeed, the idea of using research about human cognitive architecture to design instruction that facilitates learning despite the limited capacity of our working memory was described as "… the ultimate aim of the theory" (p. vii) in Sweller et al.'s (2011) extensive updated version of cognitive load theory. Moreover, Sweller et al. (2011) provided a well-grounded and thorough conceptualization of working memory, and both the common ground and the differences between a cognitive-load perspective on working memory and Baddeley's (1992, 1999, 2000) model of working memory were discussed. The differences between a cognitive-load perspective and Baddeley's model of working memory were also reflected in Schüler et al.'s (2011) review of research on working memory in multimedia instruction, in which those authors concluded that the two theories at least appear to be inconsistent. It is noteworthy that the majority of studies in this review that did provide an explicit and clear definition of working memory based this definition on Baddeley's model and not on Sweller et al.'s (2011) conceptualization of working memory.

It is also noteworthy that the procedures for measuring cognitive load were described superficially in the reviewed studies. Without an explicit and detailed description of the instructions participants received prior to problem solving or the instructional phase, it is impossible to determine whether the measurement imposed a cognitive load. That is, if participants prior to problem solving or learning are told to pay attention to task difficulty, mental effort, or cognitive load, the continuous attention directed towards this task could arguably impose an extraneous load during problem solving.

In studies using immediate ratings of cognitive load, the extraneous load imposed by the measure could potentially be even more profound, as the learner would receive a prompt to allocate attention, and thus working-memory capacity, towards monitoring cognitive load every time a cognitive-load item is administered during learning. The measurement of cognitive load could thus be conceived as a secondary task, and the accuracy of the ratings could become lower when tasks become more challenging, particularly among learners with low working memory capacity. On the other hand, if participants do not receive any instruction about the measure of cognitive load that will follow, the cognitive load of the task will be less well monitored, and participants would thus have to base their rating of cognitive load on a retrospective estimate of invested mental effort or task difficulty, a task they did not prepare for. In both cases, participants' working memory capacity must be considered.

When de Jong (2010) reviewed the 35 most frequently cited studies with "cognitive load" as a descriptor, he found that working memory was rarely measured. Our review of the most recent research in the area corroborates this finding. Based on his review, de Jong (2010) strongly recommended that working memory should be included as a control variable to ensure that differences between intervention and control groups cannot be attributed to differences in working memory, which is particularly important in quasi experimental work. In addition, measures of working memory capacity are needed to assess the extent to which the potential intrusiveness of subjective cognitive load measures may vary with working memory capacity. In this regard, a stronger relationship between ratings of cognitive load and performance among learners with higher working-memory capacity compared to participants with poorer working-memory capacity might suggest that subjective ratings of cognitive load impose an extraneous load. Yet another reason to measure working memory capacity in future research on multimedia learning framed by cognitive load theory is that such measures may be used to validate subject measures of cognitive load by examining to what extent lower ratings of cognitive load are associated with higher working memory capacity and vice versa.

Finally, operationalization and measurement of working memory should be based on a clear conceptualization of this construct. For example, a conceptualization of working memory that does not consider a central executive component but rather brings prior domain knowledge to the forefront (Sweller et al., 2011) may make it less pertinent to address executive functioning skills per se (cf., Follmer, 2018) and more pertinent to include the variable of prior domain knowledge. Attention to prior domain knowledge in the research designs is, of course, essential to investigate the expertise reversal effect and the implications of this effect for adapting instructional methods to learners at different levels of domain knowledge (Sweller et al., 2011).

**Subjective Measures of Cognitive Load**

Despite the critique of very general measures of cognitive load including few items (e.g., de Jong, 2010; Naismith et al., 1015), our review showed that variations of the Paas (1992) scale were most frequently used to measure cognitive load. One issue concerning the wide use and adaption of this scale into new contexts and cultures is the lack of psychometric validation. In their systematic review of the use of cognitive-load measures in simulation-based medical training, Naismith et al. (2015) found that the extensive use of adapted versions of the Paas (1992) scale (e.g., changing the number of points on the Likert-scale, or modifying the item to ask about perceived difficulty instead of invested mental effort) requires the reestablishment of the psychometric properties of the instrument, which was seldom performed in the research they reviewed.

While Naismith et al. (2015) examined measurements of cognitive load within simulations in medical training, de Jong (2010) found similar results in his review of studies from a broader set of educational contexts: There were several adaptions of the Paas (1992) scale, including changes in the number of points on the Likert scale, differences in the anchoring terms of the Likert scale, or differences in the wording of the items (e.g., amount of effort, how difficult, or how easy) or the timing of the measure (when it was administered during a task), without a full discussion of how such changes may affect the reliability and validity of the measure. Our results clearly show that the lack of data on the psychometric properties of subjective measures of cognitive load is still a reason for concern in contemporary research on multimedia learning.

It is also interesting that some of the reviewed studies discussed their use of general subjective measures as a limitation. For example, Hsu et al. (2015), who used the Paas (1992) scale, discussed whether the combination of subjective measures of cognitive load, challenging learning tasks, and the complex simulation technology often used in instructional design research makes it difficult for learners to estimate their cognitive load. This is because the complex simulation technology imposes some cognitive load, and it may be difficult for learners to discern how much cognitive load comes from the learning task versus the technology. Hsu et al. (2015) therefore recommended that the specificity of questionnaires is increased and that more use is made of immediate ratings of cognitive load. Similarly, Park (2015) argued that the use of subjective single-item measures, such as the Paas (1992) scale, makes it impossible to differentiate between different types of cognitive load.

Inan et al. (2015) discussed additional issues with subjective measures of cognitive load, for example whether actual cognitive load is reflected in an individual's subjective perception of mental effort. Accordingly, Inan et al. (2015) concluded that there is a need to combine subjective measures with other ways to assess cognitive load, such as a secondary task. With this recommendation as a backdrop, it is problematic that among the studies we reviewed, only 10 studies used multiple measurements of cognitive load, typically combining subjective ratings with a secondary task. While we caution against drawing conclusions from a small number of studies, it is also noteworthy that half of those 10 studies did not obtain any statistically significant correlation between their two measures of cognitive load. However, although not studying multimedia learning, Szulewski, Gegenfurtner, Howes, Sivilotti, and van Merriënboer (2017) found that a subjective measure of cognitive load (i.e., the Paas scale) correlated substantially with a physiological measure of cognitive load (i.e., pupil size).

Our review also showed that the majority of the studies operationalized cognitive load through items asking participants to rate their mental effort or perceived difficulty in relation to a task or learning phase. In approximately half of the studies, this was the only way in which cognitive load was operationalized. Mental effort, perceived difficulty, and cognitive load are clearly related, but they are not identical (Ayres & Youssef, 2008; de Jong, 2010; Schmeck et al., 2015). That is, whereas cognitive load is experienced by the learner, mental effort is exerted by the learner (de Jong, 2010). Therefore, it seems problematic that these terms are used as synonyms in the majority of the studies we reviewed. In Paas' (1992) highly cited article, for example, cognitive load was described as a construct containing two dimensions: mental effort and mental load. In relation to this conceptualization, studies operationalizing cognitive load with mental effort as their sole indicator could be said to be based on a too narrow conceptualization, whereas studies operationalizing cognitive load through a combination of indicators such as mental effort, perceived task difficulty, frustration during task, and pressure during task could be said to be based on a too broad conceptualization, calling the construct validity of such measurements into question.

**Relationships between Cognitive Load and Learning**

Our review showed that fewer than half of the studies (32 out of 73) reported findings consistent with the hypothesized relationship between cognitive load and multimedia learning, which implies that good instructional designs lead to reduced levels of cognitive load, as well as improved multimedia learning due to reduced processing demands on working memory. In other studies, however, the instructional designs that were tested either showed improvement in learning with no reduction in cognitive load or reduced cognitive load with no effect on learning. There are several possible reasons for the lack of consistency between the theoretical assumptions and the empirical findings in the majority of the reviewed studies. For example, it is possible that in some instances the observed effects on learning were caused by something other than a reduction in cognitive load, and that in some instances a reduction in cognitive load was not sufficient to bring about improvement in learning; both possibilities are problematic from a theoretical point of view.

Another possibility is that the subjective measures that were used lacked the sensitivity required to detect relevant changes in experienced cognitive load. For example, there seem to be reliability and validity issues with the widely used Paas (1992) scale. As highlighted by de Jong (2010), the evidence for claims about the psychometric properties of this measure is scarce, and it has not been thoroughly compared with other measures of cognitive load in order to establish construct validity. Of note is also that only four of the studies we reviewed presented precise and detailed information about the administration of the subjective self-report measure used, making it difficult to preclude the possibility that the administration and the procedures of the cognitive load measurements might have influenced the results. Finally, it is conceivable that the lack of consistency between hypotheses and findings in the majority of the reviewed studies in some instances might be due to the complexity and difficulty of the tasks used to measure learning and achievement. In any case, without reliable, valid, and sensitive measurements of cognitive load, the interpretation of findings becomes difficult in terms of the cognitive mechanisms underlying effects (or lack of effects) on multimedia learning. On the other hand, a more precise understanding of these cognitive mechanisms may not only allow for further theoretical validation of the core assumptions of cognitive load theory but also lead to improved instructional designs targeting these mechanisms.

## Conclusion

Cognitive load theory has been important in research on the potential benefits of using multimedia for learning. Research conducted within this perspective has driven instructional design in technology-rich educational contexts forward. Cognitive load theory has also been important in building connections between cognitive science and instructional practice. At the same time, cognitive load theory has been the subject of critique, in particular regarding measurement. The current review of research on multimedia learning framed by cognitive load theory demonstrates that there are lingering issues with the attention to working memory and the use of subjective measures of cognitive load in the most recent research in this area.

First, the vast majority of the studies we reviewed did not include a measure of working memory, nor did they provide a clear conceptualization of this core construct within cognitive load theory. This is problematic because working memory can be considered an important control variable whose inclusion also allows for the validation of subjective cognitive load measures. Second, the use of subjective measures of cognitive load was still characterized by general measures including very few items, lack of psychometric information, lack of combination with other types of cognitive load measures, and inconsistent operationalization, as well as lack of clear descriptions of the procedures used for measuring cognitive load. This is problematic because reliable, valid, and sensitive measures of cognitive load are needed to understand the cognitive mechanisms underlying the effects (or lack of effects) of instructional designs.

That said, measuring cognitive load in a reliable, valid, and sensitive way is a challenging and complex task, and some researchers within cognitive load theory have invested great effort in developing more objective measures of cognitive load and comparing them to subjective measures (Korbach, Brünken, & Park, 2018; Korbach et al., 2017; Park & Brünken, 2015). Regarding subjective measures, in particular, multi-item measures targeting different types of cognitive load have recently been developed and validated (Leppink et al., 2013; Sewell et al., 2016). These include a measure attempting to capture the sources of different types of cognitive load in complex learning tasks involving multiple documents (Cerdan, Candel, & Leppink, 2018). However, these promising avenues for cognitive load measurement were represented in the body of research that we reviewed only to a limited extent.

The implications of our review seem straightforward. Future empirical work in this area needs to include measures of working memory, not only as a control variable but also to validate subjective measures of cognitive load. For example, without attention to working memory, it is not possible to determine to what extent the potential intrusiveness of subjective measures may vary with learners' working memory capacity, which seems like an interesting and important issue. Moreover, the development and testing of reliable, valid, and sensitive measures of cognitive load need to be prioritized in this area of research, for example by further developing and testing multi-item subjective measures targeting different types of cognitive load and comparing such measures with other types of measures (e.g., physiological measures). By including reliable, valid, and sensitive measures of cognitive load in future research, more precise understanding of the cognitive mechanisms underlying the effects (or lack of effects) of instructional designs may be achieved. In turn, such understanding may have theoretical implications because it allows for further validation of assumptions within cognitive load theory, and instructional implications because it allows for more precision in designing instructional interventions.

**References**

References marked with an asterisk (*) are included in the review

*Acarturk, C., & Ozcelik, E. (2017). Secondary-task effects on learning with multimedia: An investigation through eye-movement analysis. *The Journal of Experimental Education, 85*, 126-141.

Ahuja, J., & Webster, J. (2001). Perceived disorientation: An examination of a new measure to assess web design effectiveness. *Interaction with Computers, 14*, 15-29.

Alloway, T.P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning disabilities. *European Journal of Psychological Assessment, 25*, 92-98.

Amadieu, F., Mariné, C., & Laimay, C. (2011). The attention-guiding effect and cognitive load in the comprehension of animations. *Computers in Human Behavior*, *27*, 36-40.

*Amadieu, F., Salmerón, L., Cegarra, J., Paubel, P.V., Lemarié, J., & Chevalier, A. (2015). Learning from concept mapping and hypertext: An eye tracking study. *Educational Technology & Society, 18*, 100-112.

*Andrade, J., Huang, W.-H.D., & Bohn, D.M. (2015). The impact of instructional design on college students' cognitive load and learning outcomes in a large food science and human nutrition course. *Journal of Food Science Education, 14*, 127-135.

*Ari, F., Flores, R., Inan, F.A., Cheon, J., Crooks, S.M., Paniukov, D., & Kurukay, M. (2014). The effects of verbally redundant information on student learning: An instance of reverse redundancy. *Computers & Education, 76*, 199-204.

Austin, K.A. (2009). Multimedia learning: Cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Computers & Education, 53,* 1339-1354.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*, 389-400.

Ayres, P., & Youssef, A. (2008). Investigating the influence of transistory information and motivation during instructional animations. In P.A. Kirschner, F. Prins, V. Jonker, & G. Kanselaaer (Eds.), *Proceedings of the Eight International Conference for the Learning Sciences* (Vol. 1, pp. 68-75). Utrecht, The Netherlands: ISLS.

Baddeley, A.D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A.D. (1992). Working memory. *Science, 255*, 556-559.

Baddeley, A.D. (1999). *Essentials of human memory*. Hove, UK: Psychology Press.

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Science, 4,* 417-423.

Baddeley, A.D. (2001). Is working memory still working? *American Psychologist, 56*, 851-864.

Baddeley, A.D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *Recent advances of learning and motivation* (pp. 47-89). New York: Academic Press.

Baddeley, A.D., & Logie, R.H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 22-61). Cambridge: Cambridge University Press.

Bralfish, O., Borg, G., Dornuc, S. (1972). *Perceived item-difficulty in three tests of intellectual performance capacity (Report No. 29).* Stockholm: Institute of Applied Psychology.

Brünken, R., Plass, J.L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 53-61.

Brünken, R., Plass, J.L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science, 32,* 115-132.

Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. Plass, R. Moreno, & R. Moreno (Eds.), *Cognitive load theory* (pp. 181-202). New York: Cambridge University Press.

Brünken, R., Steinbacher, S., Plass, J.L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology, 49*, 109-119.

Butcher, K.R. (2014). The multimedia principle. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 174-205). New York: Cambridge University Press.

Cantin, R.H., Gnaedinger, E.K., Gallaway, K.C., Hesson-McInnis, M.S., & Hund, A.M. (2016). Executive functions predict reading, mathematics, and theory of mind during the elementary years. *Journal of Experimental Child Psychology, 146*, 66-78.

Cerdan, R., Candel, C., & Leppink, J. (2018). Cognitive load and learning in the study of multiple documents. *Frontiers in Education, 3*: 59. doi: 10.3389/feduc.2018.00059

Cerpa, N., Chandler, P., & Sweller, J. (1996). Some conditions under which integrated computer-based training software can facilitate learning. *Journal of Educational Computing Research, 15*, 345-367.

Chan, E., & Unsworth, L. (2011). Image-language interaction in online reading environments: Challenges for students' reading comprehension. *Australian Educational Researcher, 38*, 181-202.

Chang, C.-C., Lei, H., & Tseng, J.-S. (2011). Media presentation mode, English listening comprehension, and cognitive load in ubiquitous learning environments: Modality effct or redundancy effect? *Australasian Journal of Educational Technology, 27*, 633-654.

*Chang, C.-C., Liang, C., Chou, P.-N., & Lin, G.-Y. (2017). Is game-based learning better in flow experience and various types of cognitive load than non-game based learning? Perspective from multimedia and media richness. *Computers in Human Behavior, 71,* 218-227

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8,* 293-332.

*Chen, C.M., & Wu, C.H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80,* 108-121.

*Chen, M.-P., Wang, L.-C., Chen, H.-J., & Chen, Y.-C. (2014). Effects of type of multimedia strategy on learning of Chinese characters for non-native novices. *Computers & Education, 70,* 41-52.

Chen, O., Castro-Alonso, J.C., Paas, F., & Sweller, J. (2018). Extending cognitive load theory to incorporate working memory resource depletion: Evidence form the spacing effect. *Educational Psychology Review, 30,* 483-501.

*Cheng, P.-Y., Huang, Y.-M., Shadiev, R., Hsu, C.-W., & Chu, S.-T. (2014). Investigating the effectiveness of video segmentation on decreasing learners’ cognitive load in mobile learning. *New Horizons in Web Based Learning, 8699,* 122-129.

*Cheng, T.S., Lu, Y.C., & Yang, C.H. (2015). Using the multi-display teaching system to lower cognitive load. *Educational Technology & Society, 18,* 128-140.

*Chiu, T.K.F., & Mok, I.A.C. (2017). Learner expertise and mathematics different order thinking skills in multimedia learning. *Computers & Education, 107,* 147-164.

Chuang, H.-C., & Liu, H.-C. (2012). Effects of different multimedia presentations on viewers' information-processing activities measured by eye-tracking technology. *Journal of Science Education and Technology, 21,* 276-286.

*Chung, S., Cheon, J., & Lee, K.-W. (2015). Emotion and multimedia learning: An investigation of the effects of valence and arousal on different modalities in an instructional animation. *Instructional Science, 43*, 545-559.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25,* 315-324.

*Costley, J., & Lange, C. (2017). The effects of lecture diversity on germane load. *International Review of Research in Open and Distributed Learning, 18*, 27-46.

*Craig, S.D., & Schroeder, N.L. (2017). Reconsidering the voice effect when learning from a virtual human. *Computers & Education, 114,* 193-205.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design. *Instructional Science, 38,* 105-134.

*de Oliveira Neto, J.D., Huang, W.D., & de Azevedo Melli, N.C. (2015). Online learning: Audio or text? *Educational Technology Research and Development, 63,* 555-573.

DeLeeuw, K.E., & Mayer, R.E. (2008). A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100*, 223-234.

Dillon, A., & Jobst, J., (2005). Multimedia learning with hypermedia. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 569-588). New York: Cambridge University Press.

*Dindar, M., Yurdakul, I.K., & Dönmez, F.I. (2015). Measuring cognitive load in test items: Static graphics versus animated graphics. *Journal of Computer Assisted Learning, 31*, 148-161.

Eitel, A., Kühl, T., Scheiter, K., & Gerjets, P. (2014). Disfluency meets cognitive load in multimedia learning: Does harder-to-read mean better-to-understand? *Applied Cognitive Psychology, 28*, 488-501.

Eysink, T.H., de Jong, T., Berthold, K., Kolloffel, B., Opfermann, M., & Wouters, P. (2009). Learner performance in multimedia learning arrangements: An analysis across instructional approaches. *American Educational Research Journal, 46*, 1107-1149.

*Fenesi, B., & Kim, J.A. (2014). Learners misperceive the benefits of redundant text in multimedia learning. *Frontiers in Psychology, 5*, article 710.

Follmer, D.J. (2018). Executive function and reading comprehension: A meta-analytic review. *Educational Psychologist, 53*, 42-60.

Gathercole, S.E., Pickering, S.J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessment at 7 and 14 years of age. *Applied Cognitive Psychology, 18*, 1-16.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*, 33-58.

Gerjets, P., Scheiter, K., & Cierniak, G. (2009). The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories. *Educational Psychology Review, 21*, 43-54.

Hartman, D.K., Morsink, P.M., & Zheng, J. (2010). From print to pixels: The evolution of cognitive conceptions of reading comprehension. In E.A. Baker (Ed.), *The new literacies: Multiple perspectives on research and practice* (pp. 131-164). New York: Guilford.

*Hawlitschek, A., & Joeckel, S. (2017). Increasing the effectiveness of digital education games: The effects of a learning instruction on students' learning, motivation and cognitive load. *Computers in Human Behavior, 72*, 79-86.

*Hoogerheide, V., Loyens, S.M.M., & van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior*, *41*, 80-91.

*Hsu, Y., Gao, Y., Liu, T.C., & Sweller, J. (2015). Interactions between levels of instructional detail and expertise when learning with computer simulations. *Educational Technology & Society, 18*, 113-127.

*Huang, K.-L., Chen, K.-H., & Ho, C.-H. (2014). Enhancing learning outcomes through new e-textbooks: A desirable combination of presentation methods and concept maps. *Australasian Journal of Educational Technology, 30*, 600-618.

*Huang, K., Chen, C.H., Wu, W. S., & Chen, W.U. (2015). Interactivity of question prompts and feedback on secondary students' science knowledge acquisition and cognitive load. *Educational Technology & Society, 18*, 159-171.

*Huang, X., & Mayer, R.E. (2016). Benefits of adding anxiety-reducing features to a computer-based multimedia lesson on statistics. *Computers in Human Behavior, 63,* 293-303.

*Hung, I.-C., Lin, L.-I., Fang, W.-C., & Chen, N.-S. (2014). Learning with the body: An embodiment-based learning strategy enhances performance of comprehending fundamental optics. *Interacting with Computers, 26,* 360-371.

*Inan, F.A., Crooks, S.M., Cheon, J., Ari, F., Flores, R., Kurucay, M., & Paniukov, D. (2015). The reverse modality effect: Examining student learning from interactive computer-based instruction. *British Journal of Educational Technology, 46,* 123-130.

*Izmirili, S., & Kurt, A.A. (2016). Effects of modality and pace on achievement, mental effort, and positive affect in multimedia learning environments. *Journal of Educational Computing Research, 54*, 299-325.

Jackson, S.A., & Marsh, H.W. (1996). Development and validation of a scale to measure optimal experience: The flow state scale. *Journal of Sport and Exercise Psychology, 18*, 13-35.

*Jan, J.-C., Chen, C.-C., & Huang, P.-H. (2016). Enhancement of digital reading performance by using a novel web-based collaborative reading annotation system with two quality annotation filtering mechanisms. *International Journal of Human-Computer Studies, 86*, 81-93.

*Johnson, A.M., Ozogult, G., & Reisslein, M. (2015). Supporting multimedia learning with visual signaling and animated pedagogical agent: Moderating effects of prior knowledge. *Journal of Computer Assisted Learning, 31,* 97-115.

*Johnson, A.M., Reisslein, J., & Reisslein, M. (2014). Representation sequencing in computer-based engineering education. *Computers & Education, 72,* 249-261.

*Johnson, A.M., Reisslein, J., & Reisslein, M. (2015). Transitional feedback schedules during computer-based problem-solving practice. *Computers & Education, 81,* 270-280.

*Jung, J., Kim, D., & Na, C. (2016). Effects of WOE presentation types used in pre-training on the cognitive load and comprehension of content in animation-based learning environments. *Educational Technology & Society, 19,* 75-86.

Kalyuga, S. (2000). When using sound with a text or picture is not beneficial for learning. *Australasian Journal of Educational Technology, 16,* 161-172.

Kalyuga, S. (2007). Enhancing instructional efficency on interactive e-learning environments: A cognitive-load perspective. *Educational Psychology Review, 19,* 387-399.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23,* 1-19.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors, 40,* 1-17.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351-371.

Kalyuga, S., Chandler, P., Touvinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579-588.

Kamil, M.L., & Chou, H.K. (2009). Comprehension and computer technology: Past results, current knowledge, and future promises. In S.E. Israel & G.G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 289-304). New York: Routledge.

*Kizilcec, R.F., Bailenson, J.N., & Gomez, C.J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology, 107*, 724-739.

Kok, E.M., & Jarodzka, H. (2017). Before your very eyes: The value and limitations of eye tracking in medical education. *Medical Education, 51*, 114-122.

*Korbach, A., Brünken, R., & Park, B. (2016). Learner characteristics and information processing in multimedia learning: A moderated mediation of the seductive details effect. *Learning and Individual Differences, 51*, 59-68.

*Korbach, A., Brünken, R., & Park, B. (2017). Measurement of cognitive load in multimedia learning: A comparison of different objective measures. *Instructional Science, 45,* 515-536.

Korbach, A., Brünken, R., & Park, B. (2018). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review, 30,* 503-529.

*Kozan, K., Erçetin, G., & Richardson, J.C. (2015). Input modality and working memory: Effects on second language text comprehension in a multimedia learning environment. *System, 55*, 63-73.

*Kühl, T., Eitel, A., Damnik, G., & Körndle, H. (2014). The impact of disfluency, pacing, and students‘ need for cognition on learning with multimedia. *Computers in Human Behavior, 35,* 189-198.

*Leahy, W., & Sweller, J. (2016). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science, 44,* 107-123.

*Lee, H. (2014). Measuring cognitive load with electroencephalography and self-report: Focus on the effect of English-medium learning for Korean students. *Educational Psychology, 34,* 838-848.

*Lee, H., & Mayer, R.E. (2015). Visual aids to learning in a second language: Adding redundant video to a video lecture. *Applied Cognitive Psychology, 29,* 445-454.

*Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning, 11,* 89-105.

Leppink, J., Paas, F., van der Vleuten, C.P.M., van Gog, T., & van Merriënboer, J.J.G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*, 1058-1072.

Leutner, D., Leopold, C., & Sumfleth, E. (2009). Cognitive load and science text comprehension: Effects of drawing and mentally imagining text content. *Computers in Human Behavior, 25*, 284-289.

*Liew, T.W., Zin, N.A.M., & Sahri, N. (2017). Exploring the affective, motivational, and cognitive effects of pedagogical agent enthusiasm in a multimedia learning environment. *Human-centric Computing and Information Sciences, 7:9.*

*Lin, C.-C., & Yu, Y.-C. (2016). Effects of presentation modes on mobile-assisted vocabulary learning and cognitive load. *Interactive Learning Environments, 25*, 528-542.

*Lin, J.J.H., Lee, Y.-H., Wang, D.-Y., & Lin, S.J. (2016). Reading subtitles and taking enotes while learning scientific materials in a multimedia environment: Cognitive load perspectives on EFL students. *Educational Technology & Society, 19,* 47-58.

List, A., & Alexander, P.A. (Eds.) (2018). Toward a model of multiple-source, multi-modal processing [Special issue]. *Learning and Instruction, 57*, 1-86.

Lunts, E. (2002). What does the literature say about the effectiveness of learner control in computer-assisted instruction? *Electronic Journal for the Integration of Technology in Education, 1,* 59-75.

Mahdjoubi, L., & A-Rahman, M.A. (2012). Effects of multimedia characteristics on novice CAD learners' practice performance. *Architectural Engineering and Design Management*, *8*, 214-225.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*, 49-63.

*Marefat, H., Rezaee, A.A., & Naserieh, F. (2016). Effect of computerized gloss presentation format on reading comprehension: A cognitive load perspective. *Journal of Information Technology Education: Research, 15,* 479-501.

*Mattis, K.V. (2015). Flipped classroom versus traditional textbook instruction: Assessing accuracy and mental effort at different levels of mathematical complexity. *Technology, Knowledge, and Learning, 20*, 231-248.

Mayer, R.E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81*, 240-246.

Mayer, R.E. (Ed.) (2005). *The Cambridge handbook of multimedia learning.* New York: Cambridge University Press.

Mayer, R.E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mayer, R.E. (2014). Introduction to multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 1-24). New York: Cambridge University Press.

Mayer, R.E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88,* 64-73.

Mayer, R.E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279-315). New York: Cambridge University Press.

Mayer, R.E., & Gallini, J. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*, 715-726.

Mayer, R.E., Heiser, H., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 187-198.

*Mihalca, L., Mengelkamp, C., Schnotz, W., & Paas, F. (2015). Completion problems can reduce the illusions of understanding in a computer-based learning environment on genetics. *Contemporary Educational Psychology, 41,* 157-171.

*Milenković, D.D., Segedina, M.D., & Hrin, T. N. (2014). Increasing high school students' chemistry performance and reducing cognitive load through an instructional strategy based on the interaction of multiple levels of knowledge representation. *Journal of Chemical Education, 91*, 1409-1416.

Miyake, A., & Shah, P. (1999). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442-481). New York: Cambridge University Press.

Mousavi, S.Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319-334.

Naismith, L.M., & Cavalcanti, R.B. (2015). Validity of cognitive-load measures in simulation-based training: A systematic review. *Academic Medicine, 90*, 24-35.

Naismith, L.M., Cheung, J.J.H., Ringsted, C., & Cavalcanti, R.B. (2015). Limitations of subjective cognitive-load measures in simulation-based procedural training. *Medical Education, 49*, 805-814.

*Nebel, S., Schneider, S., Beege, M., Kolda, F., Mackiewicz, V., & Rey, G.D. (2017). You cannot do this alone! Increasing task interdependence in cooperative educational videogames to encourage collaboration. *Educational Technology Research and Development, 65*, 993-1014.

Paas, F. (1992). Training strategies for attaining transfer of a problem-solving skill in statistics: A cognitive-load perspective. *Journal of Educational Psychology, 84*, 429-434.

Paas, F., Ayres, P., & Pachman, M. (2008). Assessment of cognitive load in multimedia learning: Theory, methods, and applications. In D.H. Robinson & G. Schraw (Eds.), *Recent innovations in educational technology that facilitate student learning* (pp. 11-35). Charlotte, NC: Information Age.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, *38*, 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, *32*, 1-8.

Paas, F., Tuovinen, J.E., Tabbers, H., & van Gerven, P.W.M. (2003). Cognitive load measurements as a means to advance cognitive load theory. *Educational Psychologist, 38,* 63-71.

Paas, F., & van Merriënboer, J.J.G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*, 122-133.

Paas, F., van Merriënboer, J.J.G., & Adam, J.J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419-430.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Oxford University Press.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. New York: Oxford University Press.

*Park, B., & Brünken, R. (2015). The rhythm method: A new method for measuring cognitive load – an experimental dual-task-study. *Applied Cognitive Psychology, 29*, 232-243.

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior, 44*, 267-278.

*Park, B., Korbach, A., & Brünken, R. (2015). Do learner characteristics moderate the seductive-details-effect? A cognitive-load study using eye-tracking. *Educational Technology & Society, 18*, 24-36.

*Park, B., Knörzer, L., Plass, J.L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study of the use of anthropomorphisms. *Computers & Education, 86*, 30-42.

*Park, B., Münzer, S., Seufert, T., & Brünken, R. (2016). The role of spatial ability when fostering mental animation in multimedia learning: An ATI-study. *Computers in Human Behavior*, *64*, 497-506.

*Park, S. (2015). The effects of social cue principles on cognitive load, situational interest, motivation, and achievement in pedagogical agent multimedia learning. *Educational Technology & Society, 18*, 211-229.

Pazzaglia, F., Toso, C., & Cacciamani, S. (2007). The specific involvement of verbal and visuospatial working memory in hypermedia learning. *British Journal of Educational Technology, 39,* 110-124.

Purpura, D.J., & Ganley, C.M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics. *Journal of Experimental Child Psychology, 122*, 104-121.

*Radulović, B., Stojanović, M., & Županec, V. (2016). The effects of laboratory inquire-based experiments and computer simulations on high school students' performance and cognitive load in physics teaching. *Zbornik Instituta za Pedagoska Istrazivanja, 48,* 264-283.

*Renkl, A., Skuballa, I.T., Schwonke, R., Harr, N., & Leber, J. (2015). The effects of rapid assessments and adaptive restudy prompts in multimedia learning. *Educational Technology & Society, 18,* 185-198.

Roelle, J., Lehmkuhl, N., Beyer, M.-U., & Berthold, K. (2015). The role of specificity, targeted learning activities, and prior knowledge for the effects of relevance instruction. *Journal of Educational Psychology, 107*, 705-723.

Ryu, J.H., & Yim, J.H. (2009). An exploratory validation for the constructs of cognitive load. *The Journal of Educational Information and Media, 15*(2), 1-27

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review, 19,* 285-307.

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Difference between immediate and delayed ratings. *Instructional Science, 43*, 93-114.

*Schneider, S., Nebel, S., Pradel, S., & Rey, G.D. (2015). Mind your Ps and Qs! How polite instructions affect learning with multimedia. *Computers in Human Behavior, 51*, 546-555.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*, 469-508.

Schüler, A., Scheiter, K., & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Educational Psychology Review, 23,* 389-411.

Sesma, H.W., Mahone, E.M., Levine, T., Eason, S.H., & Cutting, L.E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology, 15*, 232-246.

Sewell, J.L., Boscardin, C.K., Young, J.Q., Cate, O.T., & O'Sullivan, P.S. (2016). Measuring cognitive load during procedural skills training with colonoscopy as an exemplar. *Medical Education, 50*, 682-692.

*Shadiev, R., Hwang, W.Y., Huang, Y.M., & Liu, T.Y. (2015). The impact of supported and annotated mobile learning on achievement and cognitive load. *Educational Technology & Society, 18*, 53-69.

*Sithole, S.T.M., Chandler, P., & Abeysekera, I. (2017). Benefits of guided self-management of attention on learning accounting. *Journal of Educational Psychology, 109,* 220-232.

Skuballa, I.T., Schwonke, R., & Renkl, A. (2012). Learning from narrated animations with different support procedures: Working-memory capacity matters. *Applied Cognitive Psychology, 26*, 840-847.

*Stebner, F., Kühl, T., Höffler, T.N., Wirth, J., & Ayres, P. (2017). The role of process information in narrations while learning with animations and static pictures. *Computers & Education, 104*, 34-48.

Swaak, J., & de Jong, T. (2001). Learner vs. system control in using online support for simulation-based discovery learning. *Learning Environments Research, 4*, 217-241.

Swanson, H.L., & Alloway, T.P. (2012). Working memory, learning, and academic achievement. In K.R. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational Psychology Handbook: Vol. 1. Theories, constructs, and critical issues* (pp. 327-366). Washington, DC: American Psychological Association.

Swanson, H.L., & Jerman, O. (2007). The influence of working memory on reading growth in sub-groups of children with reading disabilities. *Journal of Experimental Child Psychology*, *96*, 249-283.

Swanson, H.L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, *100*, 343-379.

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

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*, 123-138.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

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

Szulewski, A., Gegenfurtner, A., Howes, D.W., Sivilotti, M.L.A., & van Merriënboer, J.J.G. (2017). Measuring physician cognitive load: Validity evidence for a physiologic and a psychometric tool. *Advances in Health Sciences Education*, *22*, 951-968.

Tabbers, H.J., Martens, R.L., & van Merriënboer, J.J.G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74,* 71-81.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3,* 257-287.

Underwood, G., Jebbet, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 57A,* 165-182.

van Gerven, P.W.M., Paas, F., van Merriënboer, J.J.G., & Schmidt, H.G. (2006). Modality and variability as factors in training the elderly. *Applied Cognitive Psychology, 20,* 311-320.

van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 143-156). New York: Springer.

van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measures: Evidence in favor of repeated measures. *Applied Cognitive Psychology, 26*, 833-839.

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*, 16-26.

van Merriënboer, J.J.G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147-177.

*Wang, Z., & Adesope, O. (2017). Do focused self-explanation prompts overcome seductive details? A multimedia study. *Educational Technology & Society, 20*, 47-57.

*Wang, Z., Sundararajan, N., Adesope, O.O., & Ardasheva, Y. (2017). Moderating the seductive details effect in multimedia learning with note-taking. *British Journal of Educational Technology, 48*, 1380-1389.

*Woo, J.-C. (2014). Digital game-based learning supports student motivation, cognitive success, and performance outcomes. *Educational Technology & Society, 17*, 291-307.

*Yang, H.Y. (2016). The effects of attention cueing on visualizers' multimedia learning. *Educational Technology & Society, 19*, 249-262.

*Yung, H.I., & Paas, F. (2015). Effects of computer-based visual representation on mathematics learning and cognitive load. *Educational Technology & Society, 18*, 70-77.

Table 1

*Description of all studies included in the review*

| # | Article | Subjective measure of cognitive load | Rating | Wording of item(s) | No. of items | Other measure of CL | Measure of WM | Def. of WM | Representations |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Acarturk & Ozcelik (2017) | Kalyuga et al. (1999) | D | ML | 1 | X | - | X | VW, SG |
| 2 | Amadieu et al. (2015) | Paas (1992)/Ahuja & Webster (2001) | D | ME/DT | 1/3 = 4 | - | - | - | VW, SG |
| 3 | Andrade, Huang, & Bohn (2015) | Leppink et al. (2013) | I | PC, CLM, PL | 10 (R) | - | - | - | VW, VS, SG, DG |
| 4 | Ari et al. (2014) | Paas & Van Merriënboer (1994) | D | ME | 1 | - | - | X | VW, VS, SG |
| 5 | Chang, Liang, Chou, & Lin (2017) | Adaption of Gerjets et al. (2009) and Leppink et al. (2013) | D | PC, DT, ME, CLM, PL | 12 | - | - | - | VW, SG, DG |
| 6 | Chen, Wang, Chen, & Chen (2014) | Adaption of Sweller et al. (1998) | I | DT | 4 (R) | - | - | X | VW, VS, SG, DG |
| 7 | Chen & Wu (2015) | Sweller et al. (1998) | D | ME, ML | 4 | - | - | - | VW, VS, SG, DG |
| 8 | Cheng, Huang, Shadiev, Hsu & Chu (2014) | Sweller et al. (1998)/NASA TLX | D | Not reported | Not reported | - | - | - | VS, DG |
| 9 | Cheng, Lu, & Yang (2015) | Cerpa et al. (1996) | D | ME, DT, P | 4 | - | - | - | VW, SG, DG |
| 10 | Chiu & Mok (2017) | Paas (1992) | D | ME | 1 | - | - | - | VW, SG |
| 11 | Chung, Cheon, & Lee (2015) | Paas (1992) | D | ME | 1 | - | - | X | VW, VS, DG |
| 12 | Costley & Lange (2017) | Adapted version of Leppink et al. (2013) | D | PL | 4 | - | - | - | VW, VS, SG, DG |
| 13 | Craig & Schroeder (2017) | Paas (1992) | D | ME | 1 | - | - | - | VW, VS, SG, DG |
| 14 | de Olivero Neto, Huang, & de Azevedo Melli (2015) | Paas & Van Merriënboer (1994) | D | ME | 1 | X | - | - | VW, VS, SG |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 15 | Dindar, Yurdakul, & Dönmez (2015) | Paas (1992) | I | ME | 1 (R) | X | - | X | VW, SG, DG |
| 16 | Eitel, Kühl, Scheiter, & Gerjets, (2014) | | | | | | | | |
| | *Experiment 1* | Paas (1992)/Cierniak et al. (2009) | D | ME/DT | 1/1=2 | - | - | - | VW, SG |
| | *Experiment 2* | Paas (1992)/Cierniak et al. (2009) | D | ME/DT | 1/1=2 | - | - | | VW, SG |
| | *Experiment 3* | Paas (1992)/Cierniak et al. (2009) | D | ME/DT | 1/1=2 | - | - | | VW, SG |
| | *Experiment 4* | Paas (1992)/Cierniak et al. (2009) | D | ME/DT | 1/1=2 | - | X | | VW, SG |
| 17 | Fenesi & Kim (2014)* | No reference | D | DT | 1 (R) | - | - | X | VW, VS, SG |
| 18 | Hawlitschek & Joeckel (2017) | Bratfisch et al. (1972)/Paas (1992)/Salomon (1984) | D | DT/ME/ME | 2/1/1=4 | - | - | - | NS |
| 19 | Hoogerheide, Loyens, & van Gog (2014) | | | | | | | | |
| | *Experiment 1* | Paas (1992) | I | ME | 1 (R) | - | - | X | VW, VS, SG, DG |
| | *Experiment 2* | Paas (1992) | I | ME | 1 (R) | - | - | | VW, VS, SG, DG |
| 20 | Hsu, Gao, Liu, & Sweller (2015) | Paas (1992) | D | CL | 1 | - | - | - | VW, SG |
| 21 | Huang, Chen, & Ho (2014) | Adapted from Cerpa et al. (1996) and Kalyuga (2000) | D | Not reported | 5 | - | - | - | VW, VS, SG, DG |
| 22 | Huang, Chen, Wu, & Chen (2015) | Gerjets et al. (2004) | D | ME, DT, F | 3 | - | - | - | VW, VS, DG |
| 23 | Huang & Mayer (2016) | Paas (1994) | I | ME | 1(R) | - | - | - | VW, VS, DG |
| 24 | Hung, Lin, Fang, & Chen (2014) | Amadieu et al. (2011)/Paas (1992) | D | DT/ME | 1/1=2 | - | - | - | VW, SG |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 25 | Inan et al. (2015) | Paas (1992) | D | ME | 1 | - | - | - | VW, VS, SG |
| 26 | Izmirili & Kurt (2016) | Paas (1992) | I | ME | 1 (R) | - | - | X | VW, VS, DG |
| 27 | Jan, Chen, & Huang (2016) | Marcus et al. (1996) | D | ME, ML | 4 | - | - | - | VW, SG |
| 28 | Johnson, Ozogult, & Reisslein (2015) | No reference | D | DT, EA | 8 | - | - | X | VW, VS, SG, DG |
| 29 | Johnson, Reisslein, & Reisslein (2014) | No reference | D | ME, DT | 2 | - | - | - | VW,VS, SG |
| 30 | Johnson, Reisslein, & Reisslein (2015) | No reference | D | ME, DT | 2 | - | - | - | VW, SG |
| 31 | Jung, Kim, & Na (2016) | Adaption based on Cierniak et al. (2009), Jackson & Marsh (1996), and Ryu & Yim (2009) | D | DT, EA | 9 | - | - | - | VW, VS, SG, DG |
| 32 | Kizilcec, Bailenson, & Gomez (2015) | | | | | | | | |
| | *Experiment 1* | Paas (1992) | D | ME | 1 | - | - | X | VW, VS, DG |
| | *Experiment 2* | NASA TLX | D | ME, MD | 2 | - | - | | VW, VS, DG |
| 33 | Korbach, Brünken, & Park (2016)* | Paas (1992) | D | ME, DT | 2 (R) | X | X | - | VW, SG |
| 34 | Korbach, Brünken, & Park (2017)* | Paas (1992) | D | ME, DT | 2 (R) | X | X | - | VW, SG |
| 35 | Kozan, Erçetin, & Richardson (2015) | Paas (1992) | D | ME | 1 | - | X | X | VW, VS, SG |
| 36 | Kühl, Eitel, Damnik, & Körndle (2014) | Cierniak et al, (2009)/Paas (1992) | D | DT/ME | 1/1=2 | - | X | - | VS, SG |
| 37 | Leahy & Sweller (2016) | | | | | | | | |
| | *Experiment 1* | No reference | D | DT | 1 | - | - | X | VW, VS, SG |
| | *Experiment 2* | No reference | D | DT | 1 | - | - | | VW, VS, SG |
| 38 | Lee (2014) | Bralfish et al. (1972) | D | ME, DT | 2 | X | - | X | VS, DG |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 39 | Lee & Mayer (2015) | | | | | | | | |
| | *Experiment 1* | Brünken et al. (2010) | D | ME, DT | 2 | - | - | - | VS, DG |
| | *Experiment 2* | Brünken et al. (2010) | D | ME, DT | 2 | - | - | | VS, DG |
| 40 | Lehmann, Goussios, & Seufert (2016) | Klepsch & Seufert (2012) | D | DT, MD | 9 | - | X | X | VW, SG |
| 41 | Liew, Zin, & Sahri (2017) | Kalyuga et al. (2000) | D | DT | 1 | - | - | - | VW, VS, SG, DG |
| 42 | Lin & Yu (2016) | Chang, Lei & Tseng (2011) | I | ME, ML | 2 (R) | - | - | - | VW, VS, SG |
| 43 | Lin, Lee, Wang, & Lin (2016) | Adapted version of Paas (1992) | D | ME | 4 | - | - | X | VW, VS, DG |
| 44 | Marefat, Rezaee, & Naserieh (2016) | Paas (1992) | D | ME, DT | 2 | X | - | - | VW, SG |
| 45 | Mattis (2015) | Paas (1992) | I | ME | 1 (R) | - | - | X | VW, VS, DG |
| 46 | Mihalca, Mengelkamp, Schnotz, & Paas (2015) | Ayres (2006)/Paas (1992) | I | DT/ME | 1/1 (R)=2 | - | - | - | VW, SG |
| 47 | Milenković, Segedina, & Hrin (2014) | No reference | I | ME | 1 (R) | - | - | X | VS, SG |
| 48 | Nebel et al. (2017) | Adaption from Eysink et al. (2009) | D | ME, DT | 4 | - | - | - | VW, SG, DG |
| 49 | Park (2015) | Paas (1992) | I | ME | 1 | - | - | X | VW, VS, SG |
| 50 | Park & Brünken (2015)* | | | | | | | | |
| | *Experiment 1* | Paas (1992) | D | ME | 1 (R) | X | X | - | VW, SG |
| | *Experiment 2* | Paas (1992) | D | ME | 1 (R) | X | X | | VW, SG |
| 51 | Park, Korbach, & Brünken (2015)* | Paas (1992) | D | CL | 1 (R) | - | X | - | VW, SG |
| 52 | Park, Knörzer, Plass, & Brünken (2015) | Kalyuga et al. (2000)/ Paas (1992) | D | DT/ME | 1/1=2 | - | X | - | VW, VS, SG, DG |
| 53 | Park, Münzer, Seufert, & Brünken (2016)* | Paas (1992) | D | ME | 1 (R) | - | - | - | VW, SG |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 54 | Radulović, Stojanović, & Županec (2016) | No reference | I | DT | 1 (R) | - | - | - | DG, NS |
| 55 | Renkel, Skuballa, Schwonke, Harr, & Leber (2015) | No reference | I | ME, DT | 2 (R) | - | - | - | VW, SG |
| 56 | Roelle, Lehmkuhl, Beyer, & Berthold (2015) | | | | | | | | |
| | *Experiment 1* | Swaak & de Jong (2001) | D | DT | 2 | - | - | X | VW, SG |
| 57 | Schneider, Nebel, Pradel, & Rey (2015) | Eysink et al. (2009) | D | ME, CL | 7 | - | - | - | VW, VS |
| 58 | Shadiev, Hwang, Huang, & Liu (2015) | Paas et al. (1994) | D | ME, CL | 4 | X | - | - | VW, VS, SD |
| 59 | Sithole, Chandler, & Abeysekera (2017) | Paas (1992) | I | ME | 1 (R) | - | - | X | VW, SG |
| 60 | Stebner, Kühl, Höffler, Wirth, & Ayres (2017) | | | | | | | | |
| | *Experiment 1** | Paas (1992) | D | ME | 1 (R) | - | - | - | VW, VS, SG, DG |
| | *Experiment 2** | Paas (1992) | D | ME | 1 (R) | - | - | | VW, VS, SG, DG |
| 61 | Wang & Adesope (2017) | Park, Flowerday, & Brunken (2015) | D | DT | 3 | - | - | - | VW, SG |
| 62 | Wang, Sundararajan, Adesope, & Ardasheva (2017) | Park, Flowerday, & Brunken (2015) | D | DT | 4 | - | - | - | VW, SG |
| 63 | Woo (2014) | No reference | D | DT, MD | 2 | - | - | - | VW, VS, SG, DG |
| 64 | Yung & Paas (2015) | Paas (1992) | D | ME | 1 | - | - | - | VW, SG |

*Note*. D = Delayed rating, I = Immediate rating, ME = Mental effort, CL = Cognitive load, ML = Mental load, DT = Difficulty of task, P = Pressure during task, F = Frustration, MD = Mental demand, PC = Perceived complexity, CLM = Clarity of learning material, PL = Perceived learning, EA = Expended attention, R = Repeated measurement, VW = Verbal written, VS = Verbal spoken, SG = Static graphic, DG = Dynamic graphic, NS = Not specified

*The reason these studies are categorized as involving delayed as well as repeated rating is that participants were asked to rate their cognitive load once in the middle of the learning phase and once after the last task.

Table 2
*Summarization of the results*

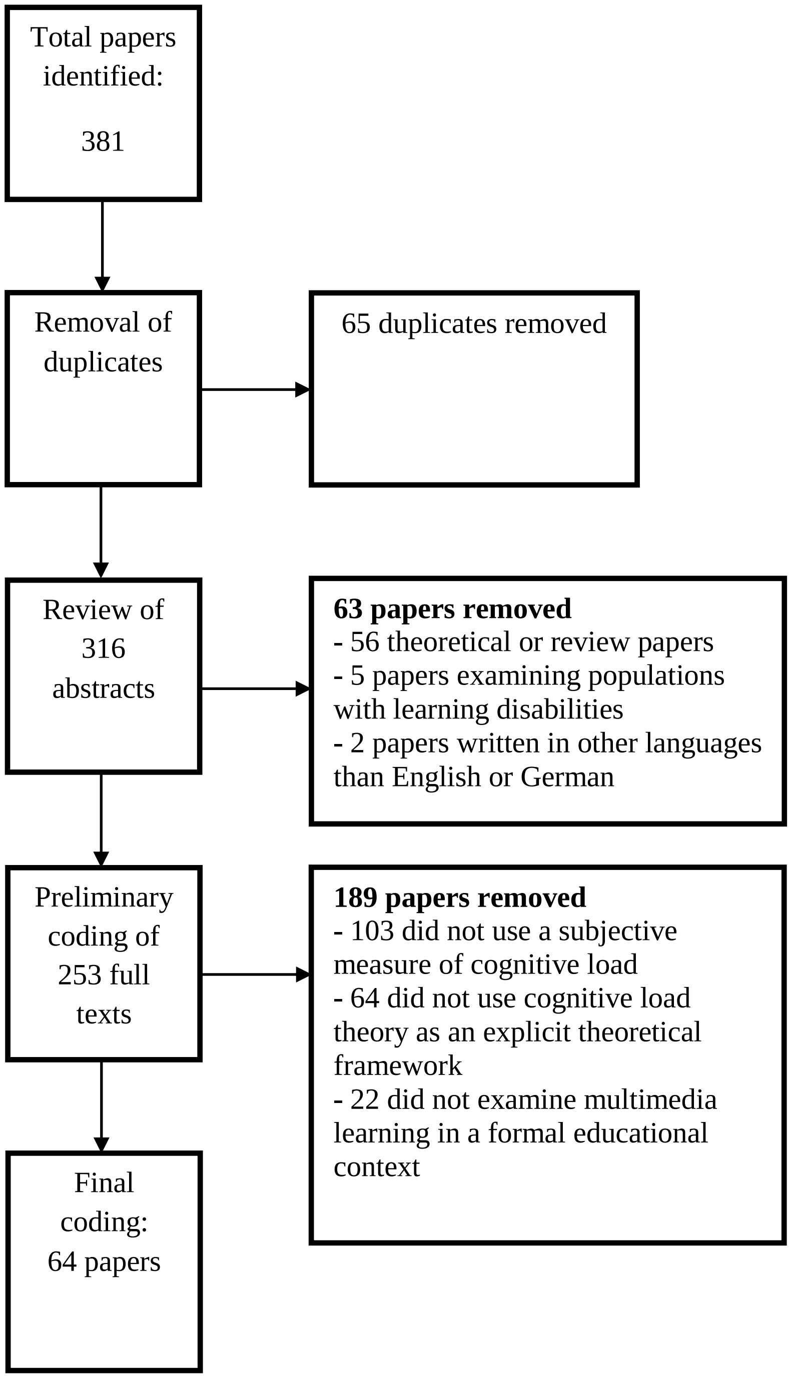
| | 2014 | 2015 | 2016 | 2017 | *N* (%) |
|---|---|---|---|---|---|
| Number of studies[1] | 17 | 29 | 13 | 14 | 73 (100) |
| Number of items in subjective measure[2] | | | | | |
| 1 item | 5 | 13 | 6 | 7 | 31 (42.47) |
| 2 items | 9 | 8 | 3 | 1 | 21 (28.77) |
| 3-4 items | 1 | 5 | 2 | 5 | 13 (17.81) |
| 5 → items | 1 | 3 | 2 | 1 | 7 (9.59) |
| Wording of items | | | | | |
| Mental effort | 12 | 24 | 8 | 9 | 53 (72.60) |
| Cognitive load | - | 4 | - | - | 4 (5.48) |
| Mental load | - | 1 | 2 | 1 | 4 (5.48) |
| Difficulty of task | 11 | 11 | 7 | 7 | 36 (49.32) |
| Pressure during task | - | 1 | - | - | 1 (1.37) |
| Frustration | - | 1 | - | - | 1 (1.37) |
| Mental demand | 1 | 1 | 1 | - | 3 (4.13) |
| Perceived learning | - | 1 | - | 2 | 3 (4.13) |
| Clarity of learning material | - | 1 | - | 1 | 2 (2.74) |
| Expended attention | - | 1 | 1 | - | 2 (2.94) |
| Perceived complexity | - | 1 | - | 1 | 2 (2.74) |
| Delayed measurement of CL | 13 | 23 | 9 | 13 | 58 (79.45) |
| Immediate measurement of CL | 4 | 6 | 4 | 1 | 15 (20.55) |
| Information about the instruction participants received about the measurement of CL prior to learning phase | - | 2 | 2 | - | 4 (5.48) |
| Additional measure of CL | 1 | 5 | 1 | 2 | 10 (13.70) |
| Provides data from reliability analysis of subjective measure of CL in own sample | 3 | 4 | 3 | 5 | 15 (20.55) |
| Explicit definition of working memory | 6 | 9 | 2 | 3 | 20 (31.25)[3] |
| Measure of working memory | 2 | 5 | 2 | 1 | 10 (13.70) |
| Expected relationship between CL and learning | 9 | 12 | 3 | 8 | 32 (43.84) |
| Participants | | | | | |
| University | 11 | 18 | 7 | 8 | 44 (60.27) |

| | | | | | |
|---|---|---|---|---|---|
| High school | 3 | 6 | 6 | 4 | 19 (26.03) |
| Elementary or middle school | 3 | 5 | - | 2 | 10 (13.70) |

*Note* [1]The number of studies are higher than the number of articles included in this review since some articles include more than one study. [2]One study does not provide information about number of items. [3]The percentage is based on number of articles, not number of studies.

Figure caption

Figure 1. *Overview of the coding process*


Total papers identified:
381
Removal of duplicates
65 duplicates removed
Review of 316 abstracts
**63 papers removed**
- 56 theoretical or review papers
- 5 papers examining populations with learning disabilities
- 2 papers written in other languages than English or German
Preliminary coding of 253 full texts
**189 papers removed**
- 103 did not use a subjective measure of cognitive load
- 64 did not use cognitive load theory as an explicit theoretical framework
- 22 did not examine multimedia learning in a formal educational context
Final coding:
64 papers