Cognitive Load Theory and Working Memory Models: Comings and Goings

Puma, Sébastien[1] & Tricot, André[2]

1. University of Cergy-Pontoise and Paragraphe laboratory (EA 349)

2. University of Toulouse 2 Jean Jaures and CLLE-LTC laboratory (CNRS UMR 5263)

To appear in S. Tindall-Ford, S. Agostinho & J. Sweller (Eds.), Advances in Cognitive Load Theory. New York: Routledge

Corresponding author details

Sébastien Puma, Paragraphe lab, University of Cergy-Pontoise, 33, boulevard du Port 95011 Cergy-Pontoise. sebastien.puma@u-cergy.fr.

Authors biographies

Sébastien Puma is a lecturer in cognitive psychology at the School of Education of Cergy-Pontoise. He was awarded his PhD (2016) from the University of Toulouse, France. His research questions focused on the involvement of working memory and attention during learning sessions; and ways to measure it. He is now part of a national research program investigating the use of Digital Games-Based Learning environment to foster arithmetic learning.

André Tricot is a professor of psychology at the School of Education, University of Toulouse. He was awarded his PhD in Cognitive Psychology (1995) from Aix-Marseille University, France. In 2014-15, he was the head of the group that designed grades 1, 2 and 3 of a new curricula for primary schools in France. André's main research topics concern the relationships between natural memory and processing (human cognitive system) and artificial memory (documents). The main question of this research is: How does designing artificial memory help natural memory instead of overloading it? Applications are in instructional design, human-computer interaction, ergonomics and transport safety.

Cognitive Load Theory and Working Memory Models: Comings and Goings

Abstract

Cognitive Load Theory uses working memory models to provide instructional guidance and to improve academic learning. Doing so, it provides empirical effects challenging existing working memory models while these models in turn provide answers and new hypotheses for Cognitive Load Theory. Yet, when challenges brought by Cognitive Load Theory stress a model beyond its capacity to handle the challenge, the Theory has to move forward to a more suitable model. The history of the relations between Cognitive Load Theory and working memory models thus cycles between empirical findings and new working memory models. The aim of this chapter is to describe this cycle before presenting emerging questions and discussing how a new working memory model could address them.

Cognitive Load Theory aims at improving academic learning by providing guidance for effective instructional design. According to this theory, when engaging in academic learning, students have to invest their resources in three types of cognitive load: intrinsic, devoted to information processing imposed by the learning task; extraneous, devoted to additional information processing imposed by the learning material presentation; and germane, devoted to learning itself, *i.e.* changing knowledge in long-term memory. Sometimes it is not possible to distinguish intrinsic and germane load. For example, when the learning task is to understand a text, the learning goal is also to understand this text.

Cognitive Load Theory uses working memory and attentional models to describe cognitive resources involved in these learning activities. It relies on existing working memory models to interpret empirical findings. Yet, some of the empirical effects unravelled by Cognitive Load Theory challenged the model used to the point where it became necessary to adopt another working memory model. The history of Cognitive Load Theory is thus a cycle between empirical findings and working memory models. The aim of this chapter is to describe this cycle before presenting emerging questions and discussing how a new working memory model could address them.

In their book on working memory models, Shah and Miyake (1999) proposed a series of criteria any working memory model should clearly address. Some of these criteria will be used to present the respective contributions of the models to the field of Cognitive Load Theory. These criteria are: “the basic mechanisms and representations in working memory”, “the unitary versus non-unitary nature of working memory”, “the nature of working memory limits”, “the relationship of working memory to long-term memory and knowledge”, and “the relationship of working memory to attention and consciousness”. Three models, namely, Baddeley’s (1986) multiple component model, Ericsson and Kintsch’s (1995) long-term working memory model and Barrouillet and Camos’ (2015) Time Based Resource Sharing model will be presented in this chapter, in a chronological order, as a means to deal with the way Cognitive Load Theory brings new challenges and questions to working memory research and how, in turn, working memory research brings answers and theoretical perspectives to Cognitive Load Theory. The challenges yet to be addressed and the limits of these relations will also be discussed.

Cognitive Load Theory before Explicit References to Working Memory Models

Cognitive Load Theory dealt with matters of limited cognitive resources (Sweller, 1988; Sweller & Levine, 1982) before using references to working memory models. It focused on how to describe the limited cognitive resources invested during learning and proposed the worked example effect and goal free effect to address the idea that students had limited cognitive resources available (*e.g.*, Broadbent, 1958; Atkinson & Shiffrin, 1968). At this early stage of Cognitive Load Theory, working memory models and cognitive architecture were not explicitly mentioned in publications.

The Worked Example Effect

One of the first findings of Cognitive Load Theory, the worked example effect was based upon the idea that cognitive resources are available for learning in limited amounts (Sweller, Ayres, & Kalyuga, 2011). Based on this idea, the worked example effect showed that novice students learned better from working on understanding the solution of a solved problem, than trying to solve the same problem on their own. Thus, freeing some of the resources that were initially dedicated to find the correct answer by providing the solution actually improved problem solving and learning outcomes. According to Cognitive Load Theory, this result was explained by a reduction of extraneous load, with a constant intrinsic load, resulting in more resources devoted to germane load. This was assumed to result in more resources dedicated to transfer information to long-term memory, in other words, in learning. At that point, there was no necessity for Cognitive Load Theory to consider working memory models, since the idea of a limited amount of cognitive resources was sufficient.

The Goal Free Effect

Another of the early Cognitive Load Theory findings, the goal free effect, is based on the same idea as the worked example effect: Freeing cognitive resources involved in problem solving helps learning (Sweller, Ayres, & Kalyuga, 2011). When novices learners try to solve a problem, they mainly use novice strategies, *e.g.* means-end analysis (Newell & Simon, 1972). This strategy, though largely transferable from one topic to another, is highly demanding. It requires learners to elaborate a problem state representation, to actively maintain the goal of the problem in working memory, and all steps between the current problem state representation and the goal state, to complete the problem solution. Cognitive Load Theory showed that removing the goal prevented students from adopting this costly strategy, freeing cognitive resources and thus improved learning outcomes (Sweller, 1988).

Initially, Cognitive Load Theory dealt with learning materials where the learning task was a problem to solve. The theory obtained improvements in learning by modifying the extraneous load: the problem solving task itself was changed, with the goal free effect and the worked example effect. Gradually, the learning materials became more complex, involving texts, graphs, sounds, animations and pictures. Cognitive Load Theory dealt with multimedia learning and, at the same time, multimedia learning researchers discovered Cognitive Load Theory. But when using multimedia materials, simple "limited short term memory resources" as described by Miller or Atkinson and Shiffrin are not sufficient. It became necessary to use working memory models to explain why cognitive load changes. Baddeley's multicomponent model became very useful to generate new hypotheses and to explain new results.

Cognitive Load Theory and the Multicomponent Model

An early working memory model used in the Cognitive Load Theory framework (Sweller, Van Merrienboer & Paas, 1998), Baddeley's multi-component model (*e.g.*, Baddeley & Hitch, 1974; Baddeley, 1986) assumed that short-term memory was also the cognitive place where information was manipulated. Changing the name, from short-term memory to working memory (Miller, Galanter, & Pribram 1960), was a change of paradigm, from a passive storage place to an active manipulation place. A major contribution of the multicomponent model of working memory was to unify what were several models of short term storage of information passing through different modality channels into a single model. This model included two slave systems, the phonological loop and the visuo-spatial sketchpad, and a central executive responsible for attention allocation. It accounted for independence between verbal and visual information retention (*e.g.* Baddeley, 1966) but also for interference in the processing of cross modal information. In this non-unitary model, units of information in the form of chunks (Miller, 1956) were maintained in two components (verbal and visual) depending on their nature. From these component, they could be used for further processing by the central executive, conceived as an attentional module, or at least responsible for the allocation of attentional resources.

In this conception, as in the conception of Cognitive Load Theory, working memory was described as the gatekeeper of long-term memory. Thus, any element to be placed in long-term memory (*i.e.* to be learned) had to pass through working memory. This description of working memory as distinct from long-term memory reflected a two way communication system, with long-term memory required when storing chunks of information. Increased long-term memory content (*i.e.* superior expertise) helped creating larger chunks. The number of chunks that could be held in working memory was limited to seven plus or minus two (Miller, 1956) but the mechanisms responsible for this limitation and forgetting in working memory were not described. Two hypotheses competed to explain this limitation and fast forgetting in working memory: interference and decay. With respect to the interference hypothesis, elements or chunks held in working memory could stay there for any duration, provided no other element entered working memory. If any other information was placed in working memory, then it would take a place and, if no place was available, would replace previous information. On the other hand, the decay hypothesis assumed that elements placed in working memory have an activation level that will decrease over time, that is, that an element's trace decays as time passes. This hypothesis assumes elements in working memory need to be refreshed to counter this time related decay and thus forgetting. The multi-component model is consistent with both hypotheses though the time related decay hypothesis might be favoured (Baddeley, 2012).

Modality Effect

The major contribution of this model to Cognitive Load Theory is probably the modality effect. The modality effect "arises when audio information replacing written text information, referring to a map, graph, diagram or tabular information results in enhanced processing and learning." (Leahy & Sweller 2016, p. 108). The multicomponent model, assuming two slave systems for maintenance and manipulation of information depending on their modalities is well suited to describe this effect. Information presented visually, such as a map, graph or diagram are maintained and processed in the visual-sketchpad while auditory information is processed in the phonological loop. This structure allows an extension of the size of working memory. The central executive is responsible for transfering information to long-term memory and to allocate attention. Thus, if information is maintained and processed in the two slave systems, its integration in a schema should be facilitated by the use of two systems instead of one. On the contrary, if information contained in the two slave systems is not consistent or not related to the same schema, this should raise the load imposed on the central executive.

Redundancy Effect

The multicomponent model was not able to explain some of the experimental results obtained in the field of Cognitive Load Theory. For example, when studying the modality effect, it was found that presenting information in two modalities could impair learning instead of improving it. More generally, presenting more information than needed has a detrimental effect on learning, imposing more information processing with no benefits for schema construction. This result was named the redundancy effect (Sweller, Ayres, & Kalyuga, 2011). A surprising finding was observed when studying redundancy: students with low levels of expertise might benefit from more information presentation, while it might have detrimental effects on more expert students' learning.

The multicomponent model, able to explain the modality effect, is not able to describe the fact that the same information presentation might have positive or negative effects on learning, depending on the expertise level of the students. This model was helpful to explain empirical results using recent learning material such as multimedia learning, when the same information was presented in different ways. A new challenge was to describe why cognitive load changed when the same information was presented in the same manner to different learners.

Cognitive Load Theory and the Long-Term Working Memory Model

The long-term working memory model (Ericsson & Kintsch, 1995) was advanced to take into account variations occurring in working memory performances associated with variations in expertise of participants with test material. In this model, working memory is a unitary phenomenon, meaning that there are no multiple modules and slave systems. It is defined as the activated part of the long-term memory, which implies that elements in working memory are representations held in long-term memory and activated through attentional processes. Knowledge in long-term memory can be viewed as composed of schemas that are larger and more complex with increasing expertise. The representations in working memory can thus be viewed as activated schemas, and they can vary with variations in expertise. Following Ericsson and Kintsch, this explains why people are merely able to recall 7 (+ / - 2) elements (Miller, 1956), or 4 in more recent estimations (Cowan, 2001) while they can remember a sentence of

20 words with little effort. More extensively than with Baddeley's multicomponent model, working memory is here viewed as relying on attentional resources, since elements (parts of long-term memory) are activated by attention. This meant giving up the spatial metaphor used in the multicomponent model, working memory being a mental place where information could be stored and manipulated (*e.g.* James, 1890; Cowan, 1988; 2014; Cowan et al. 2005). The elements held in working memory are thus defined by an activation level, which decreases over time, following a time decay hypothesis, and this decay is slower for information for which people are experts. Thus, an expert can hold large amounts of information for a longer duration at a much lower cost than a novice. What limits working memory is the level of expertise in the domain, the attentional activation of elements being dependent upon this level of expertise.

There is however no specific references to the modality of the representation of information in long-term working memory. Since it is a unitary model and all information is part of a schema, then this information should be amodal and in isolation this model is not able to explain the modality effect observed in Cognitive Load Theory research. On the other hand, it does explain the redundancy of information for expert learners presented information in two sensory modalities. If learners are expert enough, then this information presentation activates the same schema and thus imposes twice the processing for the same result.

Expertise Reversal Effect

The Long-term Working Memory model was first incorporated into Cognitive Load Theory to explain multiple empirical effects observed in research linked to expertise of the learners (Sweller, Ayres, & Kalyuga, 2011). On many occasions, instructional designs had beneficial effects on novice students learning while more advanced students showed no benefit at all or even adverse effect of the design. This model explains how, under certain circumstances, a beneficial effect might turn into redundancy. If different information activates the same schema with no direct benefits for the learner, imposing the need to process that information will impose a heavier load with no gain in schema formation, thus impairing learning.

Guidance Fading Effect

The guidance fading effect appeared when studying the worked example effect (Sweller, Ayres, & Kalyuga, 2011). A modality effect turning to a redundancy effect could be explained by the added processing for no real benefit exhibited by the expertise reversal effect. The guidance fading effect relies on a similar phenomenon to the expertise reversal effect: processing the solution of the problem has a cognitive cost with no benefit for an advanced learner and thus results in an extraneous load. Thus, after a training session on worked examples providing full guidance, it proved beneficial for learners to turn toward problem solving exercises without guidance which constitutes guidance fading.

Pace of Presentation Effect

The multicomponent model describes the benefits gained from using two different sources of information for learning, at least when these sources are consistent with each other and when learners have low levels of expertise. The long-term working memory model allows us to describe the inversion in results observed with expertise increases, for example when using the modality effect. Other studies on the modality effect found that it was also sensitive to the pacing of presentation of information (Sweller, Ayres, & Kalyuga, 2011; Ginns, 2005). The modality effect, as well as the redundancy effect, are more difficult to elicit when information pace is controlled by participants than when it is system paced. Investigating this effect, Schmidt-Weigand, Kohnert and Glowalla (2011) found that when the rhythm of presentation of information was learner paced, the benefit from dual modality presentation disappeared as compared to single modality presentation. They also found that when the presentation of information was system paced, learning outcome was sensitive to the rhythm of presentation, with a slower pace associated with a better learning performance.

Transient Information Effect

Another effect that can be related to the rhythm of information presentation is the transient information effect. The transient information effect is found when information disappears after presentation to be replaced by new information, which is the case with, for example, spoken information. If information disappears before the student has processed it, then it must be retrieved from memory or lost if not in memory. Though expertise associated with spoken language is large enough to allow large numbers of words to be maintained in working memory, this format of instruction still imposes a heavy load on working memory capacity. When considering the transient information effect, it was found to be affected by pace of presentation and segmentation. Introducing pauses during an animation or with spoken text might help learners to appropriately process information before it vanishes from working memory (Leahy & Sweller, 2011; 2016).

These phenomena of pacing of information and of transience of information do not contradict either the multicomponent or long-term working memory models, but these models are not able to describe and explain these empirical effects. Cognitive Load Theory adopted a working memory model to explain why cognitive load changed when the same information was presented in different ways. It adopted a new working memory model when confronted with the fact that cognitive load changed when the same information is presented in the same way to different learners based on their prior knowledge levels. Cognitive Load Theory now is confronted by a new challenge that may be met by adopting a new working memory model to explain why the same information, presented the same way to the same learner might impose a different cognitive load depending on time variations in presentation.

Cognitive Load Theory and the Time Based Resource Sharing Model

The Time Based Resource Sharing model (TBRS, Barrouillet, Bernardin & Camos, 2004; Barrouillet & Camos, 2015) is a recent model of working memory taking time into account to describe working memory load. It relies on four main assumptions. First, the central resource of cognition in the TBRS framework is attention (Camos & Barrouillet, 2014). Attention allows the refreshing of memory traces to prevent time related decay (Vergauwe, Dewaele, Langerock & Barrouillet, 2012). It also allows processing of information held in working memory and thus performing tasks at hand, since executive functions are deemed to rely on attention (Awh, Vogel & Oh, 2006). Second, the processes are sequential. If attentional resources are limited (*e.g.*, Broadbent, 1958), then any cognitive system has to manage them. The most widely accepted idea is that attentional resources are sequentially attributed to the different executive functions (Salvucci & Taatgen, 2010; Barrouillet & Camos, 2015 for a review). Third, the decay of memory traces is time related. There is no consensus on the nature of forgetting in working memory, some arguing that it is interference-based (Oberauer, Lewandowsky, Farrell, Jarrold & Greaves, 2012 for a review) and some arguing that it is time related (*e.g*. Baddeley, 2012; Barrouillet, Bernardin & Camos, 2004). In the TBRS model, time is considered as the main source of forgetting and chunks held active in working memory have to be refreshed periodically to prevent forgetting (see Cowan, 1995 for a similar suggestion). Since attention is the main resource in cognition and it can only be focused on one task at any time, processing a task and maintaining chunks actively requires multitasking. Fourth, multitasking can occur due to a rapid switching between processes. Attentional focus can only be applied to one chunk after another. It follows that the more chunks are kept active, the longer it takes to refresh each chunk. When chunk activation decays too much to be retrieved, it is forgotten (Barrouillet & Camos 2014). If two tasks have to be processed simultaneously, they share the attentional resource sequentially, on a time basis (see Salvucci & Taatgen, 2010 for a similar suggestion).

It follows from those assumptions that holding items active in working memory while performing a judgment or decision task will necessitate rapid switching between memory trace refreshing and information processing. From these conclusions, the authors infer that cognitive load can be modelled as a time ratio: the proportion of time spend to process information while not refreshing memory traces will have a direct impact on working memory span.

Following the above assumptions, the TBRS model allows us to predict working memory span. Working memory span is directly dependent on the amount of resources needed to process interfering tasks while maintaining and refreshing memory traces. Since one stored item can be refreshed at a time, the maximum number of elements in a memory span will be achieved when the time available to refresh memory traces is not enough to refresh all the chunks held actively. The remaining chunks, not refreshed, will be forgotten. From this, it follows that an additional task requiring time to be processed would linearly impair working memory capacity, shortening the time available for the refreshing cycle by the time needed to complete the task. If participants use strategies during long tasks (*i.e.* several seconds) to rapidly change between task processing and the refreshing cycles, short tasks requiring small amounts of attentional resources should directly impact working memory span (*e.g.* Barrouillet, Bernardin & Camos, 2004).

This suggestion has been tested several times, using a complex memory span protocol during which participants viewed letters to be memorized interleaved with distracting tasks, such as spatial judgment or mental calculus (Barrouillet *et al.*, 2007). The TBRS model allows us to infer that only the ratio between the time needed to perform the distracting task and the total time available will impact the cognitive load and therefore the working memory span. Then, using harder tasks, requiring more time to be completed, or reducing the total time available should impair working memory span by increasing cognitive load. On the contrary, while freeing attentional resources by reducing the time needed to perform a second task or increasing the total time available to complete the second task and refresh items held in working memory would result in an improvement of working memory capacity.

An interesting point inferred from the model is that the number of interfering tasks one will have to process while maintaining items active will have no effect on the memory span, since only the ratio between time needed to perform one distracting task and the time available to complete it will affect cognitive load and working memory span.

Thus, cognitive load is defined as the ratio of time spent on processing information divided by the total time allowed to complete the task. Cognitive load is then defined by the equation: $C.L. = \frac{a\,N}{T}$ with "a" being the time needed to process an interfering task, "N" the number of interfering tasks to process and "T" the available time to process the interfering tasks and refresh the memory traces. Since in a classic complex memory span protocol, participants will have N interfering tasks to process, the first equation $C.L. = \frac{a\,N}{T}$ can be transformed to: $C.L. = \frac{a\,N}{t\,N}$ with *tN* = *T*, where "t" will be the time between the onset of two consecutive interfering tasks, or the time available to complete one interfering task, and "*N*" will be the same number of interfering tasks to process. For only one processing task, cognitive load can be defined as $C.L. = \frac{a}{t}$ with "a" the time needed to process the task and "t" the time available. Considering that the time available "t" is used to both process the distracting task (*Td*) and refresh memory traces (*Tr*), cognitive load can be expressed as: $C.L. = \frac{Td}{Td+Tr}$.

The TBRS model has been used in many laboratory experiments (Barrouillet & Camos, 2015 for a review). Yet, none of these experiments have evaluated its relevance to Cognitive Load Theory. Puma, Matton, Paubel and Tricot (2018) addressed this issue. In two experiments, they used a complex span task as designed in previous studies on this model (Barrouillet *et al.*, 2007) while replacing letters to be remembered by terms of a mental calculus to be performed. Participants had thus to maintain the terms and to perform a mental calculation while performing a spatial judgment task. In their first experiment, the time ratio was manipulated by varying the difficulty of the spatial judgment task and the authors found it affected participants' performance of the arithmetic task for the less expert participants. More expert participants showed no differences in mental calculus performance in either difficulty condition, as if they were not affected by the variation of the time ratio. In a second experiments, the time ratio was varied by changing the time available to perform the spatial judgment task. Participants had to complete the same distracting task but in one condition they had one second to process each spatial judgment task while in the other condition, they had two seconds to perform the same distracting task. This resulted in a fast condition imposing a heavier load on working memory than the slow condition. Consistent with the first experiment, the results showed that novice participants were affected by the time ratio manipulation but not the experts. By varying both time allowed to perform the relevant task and the time needed or allowed to perform the concurrent task, these experiments showed an effect of time on both the intrinsic and the extraneous cognitive loads. Rather than the absolute number of elements per se, the time and the number of distracting elements should be considered when evaluating the cognitive load imposed by a given learning task.

The results of these experiments were extended in another pilot study addressing an often underlined limit of the Cognitive Load Theory: the lack of an objective measure of cognitive load (*e.g.* Paas, Tuovinen, Tabbers & Van Gerven, 2003). Electroencephalography has already been suggested as a good choice for an objective measure in Cognitive Load Theory (Antonenko, Paas, Grabner & Van Gog, 2010; Keil & Antonenko, 2017). Puma, Matton, El-Yagoubi, Paubel and Tricot (2017) used a complex span task with meaningless letters to be remembered while using the spatial judgment task used in Experiment 1 of Puma *et al.* (2018). During the entire study, they recorded the electroencephalographic data of their six participants. Results showed that the *theta* rythm (4-8Hz) varied following both the series sizes and the time ratio imposed by the spatial judgment task. Even if replication in a larger sample is needed to confirm this result, this is an argument in favor of using a model describing dynamic variations of working memory requirements for both behavioral research on instructional design and measuring cognitive load in an objective way. Using such a model might allow the discovery of new empricial effects in Cognitive Load Theory research as well as allowing explanations of already known effects, such as the pace of presentation effect.

Conclusion and Perspectives

In this chapter, working memory models were considered in relation to Cognitive Load Theory. These models contribute to explanations of empirical effects discovered using this theory. In turn, Cognitive Load Theory research questions and confronts empirical results related to these theoretical models. When these models were unable to explain empirical results provided by Cognitive Load Theory, the Theory has moved to another model more suitable to describe the range of results. The very first version of this theory only needed the general idea of limited cognitive ressources to deal with learning tasks such as problem solving. Subsequently, in order to deal with more complex materials, *e.g.* multimedia documents, Cognitive Load Theory used Baddeley's multicomponent model of working memory. Next, the long-term working memory model allowed Cognitive Load Theory to deal with learners' expertise, *i.e.* the fact that learner's previous domain specific knowledge is a key issue in instructional design. Lastly, we suggest that the Time Based Resources Sharing model allows Cognitive Load Theory to deal with another key issue of instruction: time. According to this recent working memory model, time is a cognitive resource; more accurately, this model describes memory span as the amount of resources needed to process interfering tasks while maintaining and refreshing memory traces.

With this model, we can consider extraneous load as the rate of time spent on processing irrelevant information. Therefore, when designing a learning situation, Cognitive Load Theory allows teachers to make decisions taking into account four main aspects: learning task, learning materials, learner's expertise and time. The dialogue between Cognitive Load Theory and working memory models has had a major consequence for Cognitive Load Theory: it extended its practical and theoretical scope. Future research focusing on time as a cognitive ressource, for example on the resource depletion effect (Chen, Castro-Alonso, Paas & Sweller, 2018), should confirm this major extension of Cognitive Load Theory.

References

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review*, *22*, 425-438.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation* (Volume 2), (pp. 89-195). New York: Academic Press.

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience, 139*, 201-208.

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, *18*, 302-309.

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

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *The Annual Review of Psychology*, *63*, 1-29.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *The psychology of learning and motivation*, *8*, 47-89.

Barrouillet, P., & Camos, V. (2014). On the proper reading of the TBRS model: Reply to Oberauer and Lewandowsky. *Frontiers in Psychology*, *5*, 1-3.

Barrouillet, P., & Camos, V. (2015). *Working memory: loss and reconstruction*. New York and London: Psychology Press.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 570-585.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133, 83-100.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press. Retrieved from: http://www.archive.org/details/perceptioncommunOObroa

Camos, V., & Barrouillet, P. (2014). Attentional and non-attentional systems in the maintenance of verbal information in working memory: The executive and phonological loops. *Frontiers in Human Neuroscience*, *8*, 1-11.

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

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, *104*, 163–191.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87-185

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, *26*, 197-223.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*, 42-100.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, 15, 313- 331.

James, W. (1890). *Principles of psychology*. Volume 1(Chap 11). Global Grey (204). Retrieved from: www.globalgrey.co.uk.

Keil, A. & Antonenko, P. D. (2017). Assessing working memory dynamics with electroencephalography implications for research on cognitive load. In R. Zheng (Ed.), *Cognitive load measurement and application: A theoretical framework for meaningful research and practice*. (pp. 107-125). New York: Routledge.

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology*, *25*, 943-951

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

Miller, G. A. (1956). The magical number of seven, plus or minus two. some limits on our capacity for processing information. *Psychological Review*, *101*, 343-352.

Miller, G. A., Galanter, E. & Pribram, K. H. (1960). *Plans and the structure of behaviour*. New York: Henry Holt.

Newell, A. & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice Hall.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C. & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*, 779-819.

Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*, 63-72.

Puma, S., Matton, N., Paubel, P.-V., El-Yagoubi, R. & Tricot, A. (2017). Time Based Resource Sharing model as a mean to improve cognitive load measurement. *10th International Cognitive Load Theory Conference*, November 20-22. University of Wollongong, Australia.

Puma, S., Matton, N., Paubel, P.-V. & Tricot, A. (2018). Cognitive load theory and time considerations: Using the time-based resource sharing model. *Educational Psychology Review*, *30*, 1199-1214.

Salvucci, D. D., & Taatgen, N. A. (2010). *The multitasking mind*. Oxford: Oxford University Press.

Schmidt-Weigand, F. Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction*, *20*, 100-110.

Shah, P., &Miyake, A. (1999). Models of working memory. An introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 1-27). Cambridge, UK: Cambridge University Press.

Sweller, J. & Levine, M. (1982). Effects of goal specificity on means-end analysis and learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *8*, 463-474.

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

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

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

Vergauwe, E., Dewaele, N., Langerock, N., & Barrouillet, P. (2012). Evidence for a central pool of general resources in working memory. *Journal of Cognitive Psychology*, *24*, 359-366.