Educational Psychology Review
https://doi.org/10.1007/s10648-019-09461-9

REVIEW ARTICLE

Cognitive Load Theory and Human Movement: Towards an Integrated Model of Working Memory

Stoo Sepp[1] · **Steven J. Howard**[1] · **Sharon Tindall-Ford**[1] · **Shirley Agostinho**[1] · **Fred Paas**[1,2]

Published online: 05 February 2019

Abstract

Cognitive load theory (CLT) applies what is known about human cognitive architecture to the study of learning and instruction, to generate insights into the characteristics and conditions of effective instruction and learning. Recent developments in CLT suggest that the human motor system plays an important role in cognition and learning; however, it is unclear whether models of working memory (WM) that are typically espoused by CLT researchers can reconcile these novel findings. For instance, often-cited WM models envision separate information processing systems—such as Baddeley and Hitch's (1974) multicomponent model of WM—as a means to interpret modality-specific findings, although possible interactions with the human motor system remain under-explained. In this article, we examine the viability of these models to theoretically integrate recent research findings regarding the human motor system, as well as their ability to explain established CLT effects and other findings. We argue, it is important to explore alternate models of WM that focus on a single and integrated control of attention system that is applied to visual, phonological, embodied, and other sensory and nonsensory information. An integrated model such as this may better account for individual differences in experience and expertise and, parsimoniously, explain both recent and historical CLT findings across domains. To advance this aim, we propose an integrated model of WM that envisions a common and finite attentional resource that can be distributed across multiple modalities. How attention is mobilized and distributed across domains is interdependent, co-reinforcing, and ever-changing based on learners' prior experience and their immediate cognitive demands. As a consequence, the distribution of attentional focus and WM resources will vary across individuals and tasks, depending on the nature of the specific task being performed; the neurological, developmental, and experiential abilities of the individual; and the current availability of internal and external cognitive resources.

Keywords Cognitive load theory · Working memory model · Attention · Human movement · Gesturing · Learning

✉ Stoo Sepp
Stoo@uow.edu.au

Extended author information available on the last page of the article

Introduction

Working memory (WM) theories and models, the measurement of WM capacity, and the neurological structures involved in WM have been investigated from a variety of perspectives, including experimental psychology, educational psychology, linguistics, neurology, and developmental psychology, spanning nearly 50 years of research. Within the field of education, WM insights have informed educational practice through research into attentional factors, schema acquisition and construction, spatial awareness, visual representation, learning material design, and more. Cognitive load theory (CLT), in particular, is responsible for a number of these advances through its consideration of the implications of human cognitive architecture for the characteristics and conditions for effective learning and instruction. Research in CLT has already identified a number of empirically supported effects that have led to the identification of several instructional principles that can inform teaching practice and the design of learning materials. This includes findings that information is acquired more effectively when it is integrated rather than distributed (split-attention effect; Chandler and Sweller 1992), when it is augmented multimodally (modality effect; Tindall-Ford et al. 1997), and when this information is complementary rather than redundant (redundancy effect; Kalyuga et al. 1999). It has also been found that many of these effects tend to reverse at higher levels of expertise (expertise reversal effect; Kalyuga, Ayers, Chandler, and Sweller, 2003). More recently, a number of studies investigating the human motor system's role in the learning process have suggested that gestures and other human movements can also be beneficial in educational settings. As just one example, there is now emerging evidence that physically enacting the concepts to be learned may support the acquisition of information better than simple auditory presentation (Agostinho et al. 2015; Hu et al. 2014; Mavilidi et al. 2015). These studies are typically interpreted in relation to widely accepted WM models, such as those by Baddeley and Hitch (1974), Baddeley (1986), Baddeley 2000, Baddeley 2003), and Cowan (1988, 1995, 2001).

While these WM models have previously served as a strong framework for CLT findings (Sweller 1988; Sweller and Chandler 1991; Sweller et al. 1998) and related areas, it is unclear how these theories can account for the cognitive effects of human movement deriving from CLT. For instance, within Baddeley and Hitch's (1974) multicomponent model of WM, it is unclear how movement can be integrated into its auditory, visual-spatial, and executive control systems.

In an attempt to reconcile emerging CLT insights with the principles of existing WM models, this paper will first provide an overview of central concepts related to WM. Based on findings related to WM depletion, attentional inclinations, and attentional distribution across modalities, the role of attention in WM theory is examined, with a focus on how attentional control plays an important role in cognition and learning. To account for these attentional factors, a distributed attention model of WM is proposed—one which builds upon existing models and posits a more dynamic and multidimensional approach to understanding WM processes. A closer examination of the model follows, including how it may explain well-established CLT effects. Recent research into the human motor system in educational contexts is then critically discussed in terms of how gestures may be integrated into Baddeley and Hitch's model, with the suggestion that when contrasted with the proposed model, gestures and human movements may constitute an additional modality. Finally, limitations of the proposed model are discussed, along with potential areas of future inquiry.

Models of Working Memory

WM is a theoretical construct for describing the processes and systems related to the task-relevant activation (bringing to mind), maintenance (holding in mind), and processing of mental information during the performance of a task (Baddeley and Hitch 1974). Miller (1956) originally proposed that unaided individuals' capacity to receive, remember, and process information was limited to seven units of information, plus or minus two (a figure that has since been debated as being as low as four, or even one, unit; e.g., Cowan 2010). This idea of a limited active memory capacity has served as the basic assumption for many years.

After Miller (1956), the next major advancement in WM theorizing was Atkinson and Shiffrin's (1968) model, which suggested a separation of the human memory system into three subsystems: a (1) sensory register that processes substantial sensory information for a very brief amount of time, with only the most pertinent information being temporarily activated in the (2) short-term store (STS; now analogous to short-term memory (STM)), which, if processed effectively, would be encoded into a (3) long-term store (LTS). This model is important for two reasons. First, it separated a passive receipt of sensory input (in the sensory register) from post-sensory processing (in STS), thereby positing a cognitive process that influences what becomes activated in the STS—or, more accurately, within the focus of attention—for processing. Second, it differentiated between different modalities, such that visual linguistic inputs from the sensory register (e.g., a written word) may be encoded in the STS in a corresponding visual-linguistic form.

Building upon this theorizing of modalities within the human memory system, Baddeley and Hitch's (1974) resultant multicomponent model of working memory (Baddeley 1983, 1986, 2000) sought to account for findings that extended beyond temporary activation of information in the STS, to also include processing of the information through "reasoning, comprehension and learning" (Baddeley and Hitch 1974, p. 201). While positioning their model as describing a WM system, they retained a separation of visual and auditory modalities. This multicomponent model reconceptualized STM as more than just a temporary storage mechanism, but rather as a system capable of processing information across multiple sensory inputs (a system more deserving of its description as *working* memory). As much of the research at that time involved participants completing visual and auditory retention span tasks, Baddeley and Hitch's model postulated one WM system for each, supported by a central processing system called the central executive (see Fig. 1). One of these modality-specific "slave" systems was the phonological loop, in which sound- and speech-based information is processed (Baddeley 1992). It can also register written information when processed as self-speak. The other, the visuospatial sketchpad, is responsible for processing visual and spatial information to support motor control (Baddeley and Lieberman 1980). Integrating and processing information within the slave systems is the central executive, which also serves to manage the distribution of attention.

Baddeley (2000) later added a third slave system called the "episodic buffer" (see Fig. 2), as the existing model could not account for the formation of unitary representations comprised of integrated information that spanned visual and auditory modalities. The episodic buffer was thus described as a system that integrates visual, spatial, and verbal information, while also acting as a bridge to long-term memory, facilitating retrieval from this system for integration with available task-relevant information in other WM systems (Baddeley 2000).

Cowan (1988) found that existing explanations of how information is processed immediately post-stimulus in the "pipeline" of human memory (e.g., Broadbent, 1958, Broadbent 1982) may not

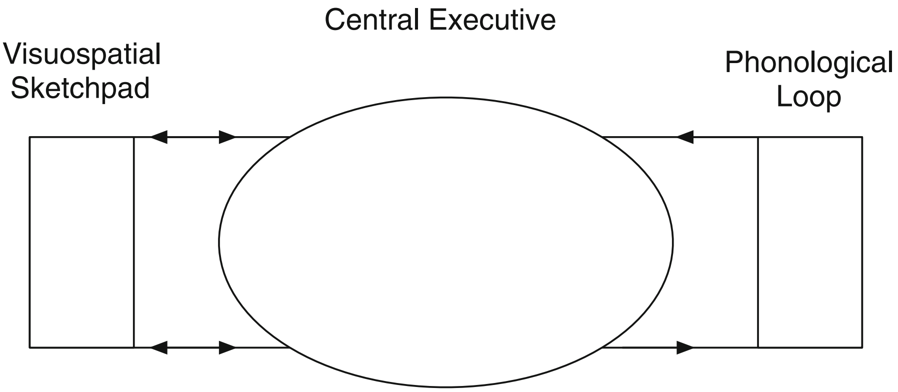


Fig. 1 Diagram based on Baddeley and Hitch's working memory model (Baddeley and Hitch 1974)

be well suited to account for the timing and sequence of mental processing across all contexts. In an attempt to reconcile these findings, he aimed to describe processes underlying the selection of certain stimuli and the dismissal of others via the application and focus of attention. Cowan's theorizing about WM included a number of departures from previous theories. For one, whereas Baddeley described different WM systems for dealing with verbal and visual-spatial information, Cowan's work posited a single central attentional resource that was responsible for dealing with all forms of information to be processed. Cowan's research also included a quantification of WM capacity, for which he posited that storage capacity (the amount of information that could be maintained in mind without rehearsal or chunking strategies) and control of attention (processes that enable the processing, combination, and reconciliation of this mentally represented information) needed to be considered separately. As such, his estimate of WM capacity as three to five items reflects the number of unrehearsed and uncombined pieces of information that can be concurrently activated in WM (Cowan, 2000, Cowan 2010).

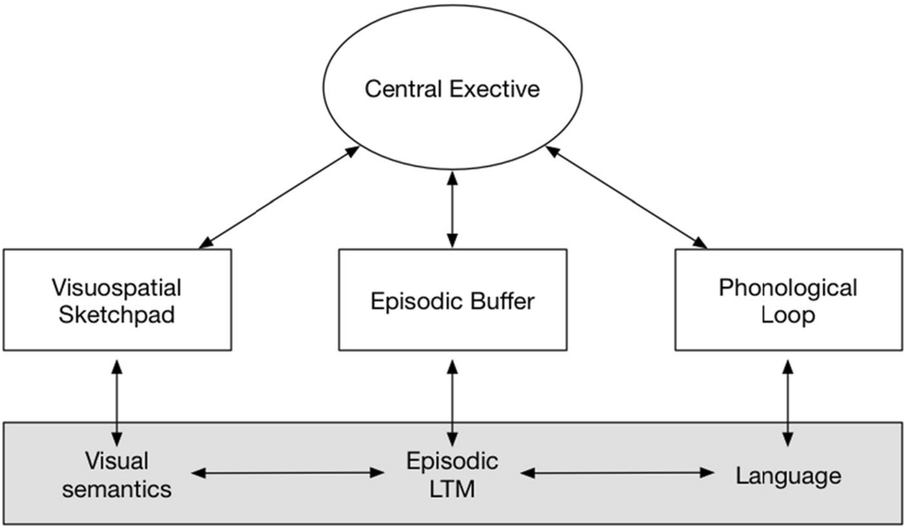


Fig. 2 Diagram based on Baddeley and Hitch's working memory model (Baddeley 2000)

In this section, an overview or historical conceptions of WM have been discussed, including the well-established models of Baddeley and Hitch (1974) and of Cowan (1998), 1999). Despite the differences between these models, the basic theoretical assumption of a limited WM capacity has persisted since Miller (1956) first reported his findings. Yet, while both theories retain their proponents and evidence base, the focus on a single attentional resource as described by Cowan has been expanded upon by a number of WM theorists. The next section discusses how the relationship between prior experience, expertise, gesturing, and attentional distribution may provide a robust foundation for reconceptualizing WM processes.

The Distribution of Attention Within a Single Integrated Working Memory System

Cowan's (1995), 1999) WM model differs significantly from Baddeley and Hitch's in that, like earlier proposals of Pascual-Leone (1970) and later formulations by Engle (2002) and Heitz and Engle (2007), there is no emphasis on distinguishing between processing different modalities. Instead, words, images, gestures, or other modalities are not relegated to a specific subsystem; rather, all information is processed through a single system with a limited capacity. According to Cowan, WM involves more than just the mental activation and maintenance of information but also, if nomenclature is taken as intended, working with information immediately following exposure to a stimulus through effortless (e.g., recoiling your hand upon touching a hot surface) and effortful attentional control processes (e.g., activating schemas after being presented with a formula on a whiteboard). A number of studies have shown that scope and control of attention are intrinsically linked in WM processes (Conway et al. 2005; Engle 2002; Kane et al. 2001), although Cowan asserts they are separable and individually constrained. For this reason, Cowan's (2000) review asserted that Miller's (1956) original "7 plus or minus 2" capacity limit was merely an estimate, and that its true limit may be around three to five items if chunking and rehearsal are restricted. Cowan's model, though important for highlighting the concept of attentional focus as a means to discuss WM processes, does not directly speak to how attention can play a role in information processing during problem solving and learning—a core focus for CLT.

To explore the relationship between attention and the processing of information during learning, an alternate conception of WM may provide a more meaningful framework for CLT and educational contexts. For instance, Engle (2002) stands in contrast to Cowan by proposing that WM capacity is not defined by individual differences in storage capacity alone, but also as a function of the ability to control attention for goal-orientated means (Engle 2002). Unsworth and Engle (2007) assert that WM capacity can actually be thought of as a product of two concurrent processes: attentional focus on task-relevant information and ignoring task-irrelevant information (suppression), and simultaneously using cues (e.g., stimuli) to reconcile information from the environment and long-term memory for task completion. Further, in contrast to Baddeley and Hitch's focus on modality-specific WM stores, this theory conceptualizes attention as a domain-free resource for information processing (which underpins WM capacity), and it diverges from Cowan in its inextricable inclusion of the processing of (working with) information within the focus of WM.

When considering CLT's *redundancy effect* though this lens, for example, as a learner focuses on two or more sources of redundant information, limited attentional resources must, regardless of the modality of information, be devoted to the reconciling and disregarding of any extraneous information, and the processing of relevant information. Successful task completion in this case would depend on the ability of each learner to use attentional control to reduce interference between redundant sources of information. Given that CLT effects such as split attention, modality, and redundancy effects often require participants to use attentional control in an attempt to establish the relationships between different sources of information, Engle's conceptualization of WM processing aligns well with individual differences in the ability to control attentional resources. If we then consider WM capacity as dependent on the ability to control a domain-free attentional resource, individual differences between learners may be explained by experiential factors such as prior learning experiences or exposure to different modalities. Differences in expertise may thus explain differential performance across visual-spatial and auditory domains, as well as their differential susceptibility to interference. For example, a musician might have more experience and thus more robust schemas and strategies (e.g., chunking, understanding viable and less plausible connections) to deal with auditory pattern information. Rather than separate WM systems, expert strategies for mobilizing and applying attention to select, process, and reconcile information in WM may explain their advantage in this domain. Indeed, de Groot's (1965) seminal study found that expertise in playing chess was determined by continual exposure, practice, and rehearsal, further suggesting that these players have more experience to deal with visual-spatial pattern information in this context (rather than a superior WM capacity, or even a generally more-advanced visual-spatial sketchpad).

When also taking into account findings that demonstrate the benefit of iconic gesturing for hearing participants, in contrast to the effects demonstrated with deaf learners, experience-related biases in the application of attention (and competence in this application across domains) may exist for all learners but merely are weighted differently. This distribution of attention would be based on an individual's previous experiences, expertise, available WM resources, as well as the nature of the immediate task at hand. For example, learners with more experience in reading may have an attentional inclination towards visuospatial information (e.g., seek, be more familiar with, and have better strategies to interpret and integrate), while learners with more experience in using their hands may have an attentional inclination towards an embodied modality. This does not mean that attentional focus is placed on a single modality at any given time, or that there are separate WM systems devoted to each, but rather an attentional distribution that constantly shifts and adjusts across modalities in a manner most efficient for the individual given the current task.

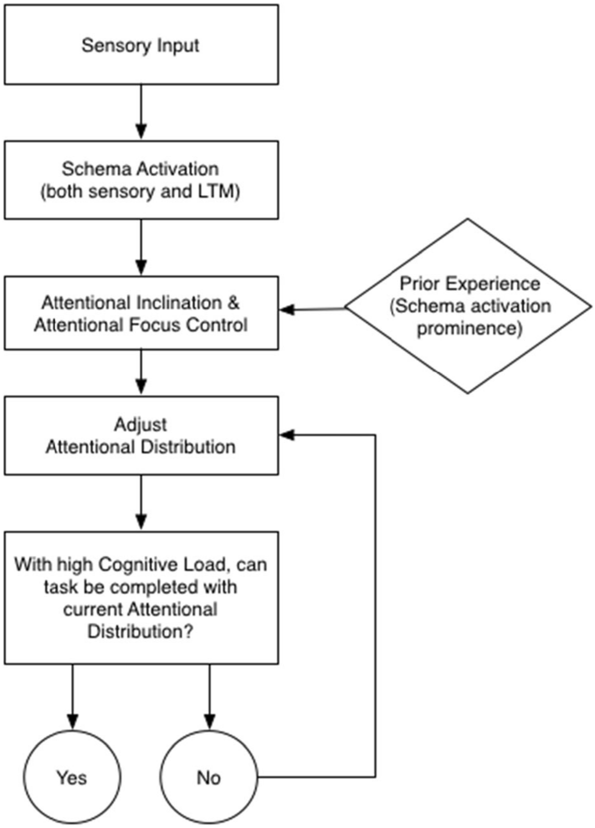
It is also important to emphasize that schemas activated, maintained, and processed across these modalities do not require attentional focus to rapidly shift between systems or modalities to be processed centrally. Instead, information across modalities can be activated, consolidated, and processed within the focus of effortful mental attention (see Pascual-Leone, 1970). This aligns with Halford et al.' (1998) work on the nature of complexity, in that complexity—regardless of the modalities from which it derives—is defined by the processing of relationships between elements (i.e., items, elements, memory traces, schemas) activated concurrently in WM plus the serial (in-sequence) processes required to reconcile them. In these formulations, there is no need to define visuospatial, auditory, or embodied (or any additional) WM subsystems. In fact, the presence of an as-yet-unknown quantity of modalities should be assumed, so as to account for future findings.

For those with less competence processing information in a particular form (e.g., written information), supportive strategies may reduce the cognitive load placed upon these learners. In line with suggestions that gestures may temporarily "offload" mentally represented information, Chu et al. (2013) found that participants with a lower visuospatial WM capacity gestured more frequently than those with higher capacities. Brucker et al. (2015) similarly found that participants with lower visuospatial abilities benefited from observing aligned gestures when learning about fish locomotion patterns. Pouw et al. (2016) found that gesturing reduced saccadic eye movements (a measure of cognitive processing) in problem-solving tasks when participants had lower visual WM capacity. This suggests that gesturing may serve as a general support mechanism for WM and attentional resources—especially if WM resources devoted to that information is cognitively taxing—without the need for reference to distinct WM systems. Indeed, Pouw et al. (2014a), in a review of the cognitive function of gestures, assert that gestures provide an external support mechanism for internal cognitive processes provided the gestures do not create a significant cognitive burden to interfere with the current task at hand (see also Pouw et al. 2014b). Schmalenbach et al. (2017) extend this argument to suggest gestures act as a supportive and compensatory mechanism when cognitive load is high, when perceptual and cognitive abilities are low, or both.

This suggests that gestures can support WM processing in two ways. First, gestures may temporarily offload mental information (Cook et al. 2012; Goldin-Meadow et al. 2001; Ping and Goldin-Meadow 2010; for a review, see Risko and Gilbert 2016), to postpone the temporal decay of a memory trace (i.e., "hold it in your hands") and remove its demand from WM. With the gesture physically maintaining this information, attentional resources normally devoted to the internal maintenance of that information are offloaded, freeing up WM resources and permitting attentional focus to be directed where it is most needed. Second, gestures can support processing in WM through the physical embodiment of a process, concept, or object. This can reframe a problem to be solved in an alternate modality, essentially externalizing the problem itself. Rather than freeing WM resources for allocation elsewhere, this serves to focus attention towards an alternative, aligned modality as a means to increase efficiency of problem solving. For example, Chu and Kita (2008) found that when participants were asked to solve mental rotation tasks, they used their hands as proxy objects for the item to be rotated. Further, Pouw et al. (2014a) assert that for the high WM demands in novel tasks, gestures can serve as a cognitive support mechanism (see Fig. 3). In contrast, distracting or redundant gestures may have the opposite effect, given both forms of information would need to be reconciled and integrated; indeed, De Koning and Tabbers (2013) found that making pointing gestures simultaneously with an animated arrow did not provide any learning benefits, a result that has implications for modality and human movement effects within CLT. By making or observing unrelated or unnecessary gestures, this may impair the processing of activated schemas and divert attentional focus from information essential for learning.

A recent development in CLT exploring the concept of WM depletion as a construct (Chen et al. 2017; Healy, Hasher, & Danilova, 2011; Schmeichel 2007) may also have implications for both attentional distribution and the use of gestures. Well-established results on the *spacing effect* demonstrate that learning episodes over time result in increased retention when compared to a single "massed" learning episode (Delaney et al. 2010; Ebbinghaus 1885/1964; Gluckman et al. 2014; Kapler et al. 2015). This *depletion effect* asserts that WM resources are not fixed and deplete over time based on varying cognitive demands—like a car's fuel gauge decreasing at different rates during a long drive across plains and through mountain passes. While this topic continues to be debated, it nevertheless supports a single attentional resource applied to the various foci of learning and suggests that gesturing may be able to forestall depletion of WM resources. It does so not by refilling the tank, so to speak, but by forestalling depletion by offloading some of the demand to a secondary resource (like a battery activating

Fig. 3 Attentional redistribution based on individual attentional inclination and cognitive demands



in a hybrid car to reduce the car's fuel consumption). Within the context of CLT, a model that attempts to reconcile across these various findings and effects has not previously been undertaken. As such, the next section will present an integrated model of WM, the one that, envisioning a single attentional resource for the activation, maintenance, and processing of information in WM, explains gestures, prior experience, expertise, and the distribution of attentional focus towards learning and problem solving.

A Distributed Attention Model of Working Memory

The proposed integration of the variety of aforementioned WM principles and applications to the breadth of CLT effects is presented in Fig. 4 as a visual representation that illustrates how multiple modalities may combine within an integrated WM system in which a common attentional resource activates all forms of information to facilitate processing. On the outside are external environmental stimuli that learners may encounter, including instruction, learning materials, and other aspects of the classroom environment or learning situation that may catch their attention. Additionally, the external environment can also include externalized supports as a means to redistribute cognitive resources in a manner more efficient for problem solving (e.g., gesturing and reading aloud).

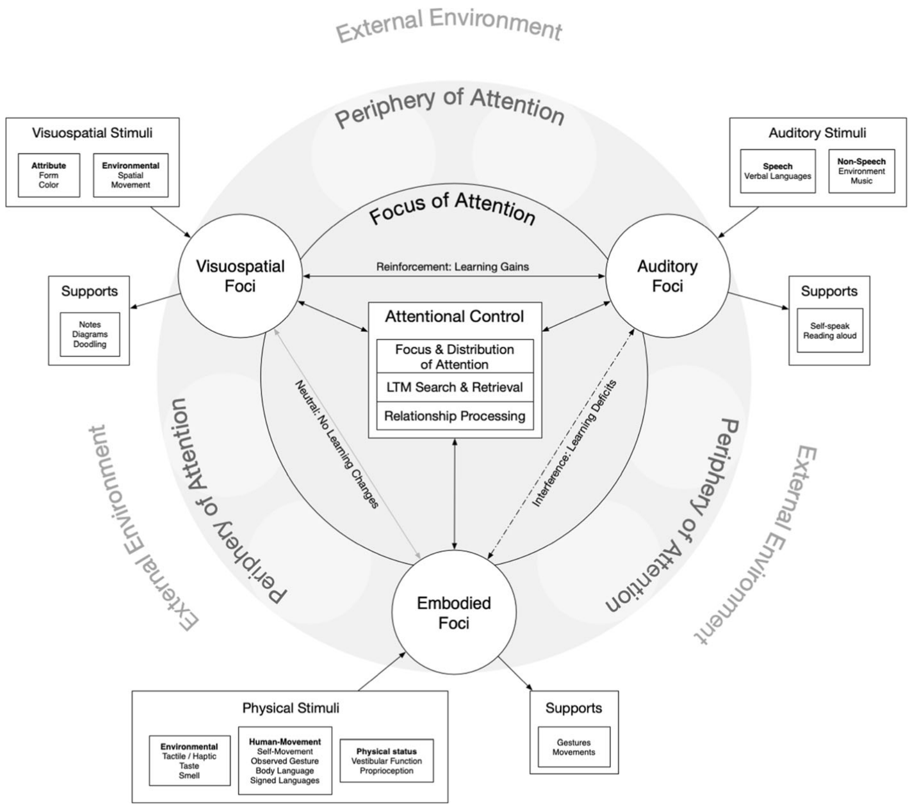


Fig. 4 The proposed distributed attention model of working memory including three traditionally investigated foci of attention

Moving inward, the large circle—encompassing the periphery of attention and the focus of attention—represents everything in WM, although the degrees of activation may vary. Following the model of Pascual-Leone (1970), the outer ring reflects information that is less activated within the periphery of attention (e.g., automatically or effortlessly), such as over-learned, novel, salient, or misleading stimuli that nevertheless capture at least some of the attentional resource. In the center circle is information within the effortful and intentional focus of attention—those things currently and intentionally being processed in mind. Following the proponents of a domain-free attentional resource that underpins WM, the focus of attention is comprised of information derived from any number of domains; visuospatial, auditory, and embodied foci are some examples. It is important to also note that foci can be both derived from the environment or searched for and activated in long-term memory (LTM). The term "foci" is chosen specifically to both include modalities and build upon them, as well as to include innumerable additional foci (e.g., sociocultural knowledge, emotional state, language proficiency, cognitive differences, with any as-yet-unidentified foci represented by blank circles). This model has included separated foci to allow for a specific exploration of each to investigate whether learning effects may apply uniquely or manifest differently across them. Foci provide opportunities to also explore how different combinations of information across them may affect reinforcement and interference and, by extension, learning.

At the center of this model, which builds upon the work of Engle (2002) and many others (Kane et al. 2001; Pascual-Leone and Baillargeon 1994; Pascual-Leone and Smith 1969; Turner and Engle 1989; Unsworth and Engle, 2007), is the effortful and intentional focus of a unitary and capacity-constrained attentional control mechanism. The limits of this system are the reason we have a WM capacity limit. Our application and control of this attentional resource—through the competitive process of our schemas and environmental cues vying for our attention—determines which foci are selected for processing and which foci are dismissed, based on the task at hand. This includes those involuntary external stimuli that successfully capture our attention (e.g., an animated or novel stimulus), as well as the effortful focus of attention on aspects of the situation we deem to be important. This information derived from the environment is also reconciled, via attentional control processes, with information activated from LTM (which is also brought into the focus of attention for recombination, integration, or reconciliation, as needed). Individuals' attentional inclinations also influence the distribution of attention focus. For example, when faced with solving a math problem in the classroom, some learners will focus attention on the voice of their instructor. Other learners may focus solely on the whiteboard, while ignoring the voice of the instructor. Still others may focus their attention on their textbook to derive the necessary information, returning their attention to the instructor only at key junctures. Each of these can be seen as the consequence of a competitive process between individual self-conceptions (i.e., beliefs about how one learns most effectively), situational cues (e.g., instructions and emphases of the facilitator), and individual schemas (e.g., learned strategies for which behaviors and processes are needed to achieve one's goals in this context). Therefore, attentional inclination towards specific modalities is based on an individual's experience and expertise, combined with the ability to control the focus of attention across those foci, which will result in a diversity of learning strategies across all students in a single classroom.

Attentional foci in different modalities are not processed independently, as articulated in dual-processing models but interdependently in the focus of attention, connecting relevant foci. The types of lines between each circle (dashed, solid, and gray) represent the different ways in which attentional control can combine foci, which may result in either interference or reinforcement. For example, when a learner engages with audiovisual learning materials in the classroom and the *modality effect* is observed, visuospatial and auditory foci are combined to reinforce concepts across sounds heard and diagrams presented. On the other hand, if auditory information conflicts with visual information, additional attentional resources are devoted to noticing and reconciling this information, resulting in interference. It should also be noted that attentional focus across these foci imply a spatial element within each, as visual, embodied, and auditory inputs can each provide spatial information to aid in situating an individual in a specific location (Rhodes 1987; Smith et al. 2009).

In the classroom example above, if the instructor walks across the room to help a student, this may redistribute WM resources for a moment through automatic and effortless shifts in attentional focus (i.e., distraction). The speed and weight of this redistribution would depend on an individual's expectancies, needs, and previous experiences. To account for this, the model assumes that attentional distribution across foci fluctuates over time, based on the attentional inclination, expertise, previous experience, and current cognitive demands of the learner. An individual's experience with particular modalities may differ, so their attentional inclination for, and shift in attentional focus towards, other modalities may also differ. This model is designed to account for this so as to present the state of a learner in multiple ways. Figure 4 presents a visualization that assumes equal attentional distribution across foci, while

Fig. 5a, b provides potential examples of how the model may conceivably display attentional inclination towards specific foci over others (with or without combination effects). Another possible static version of the model may visualize attentional distributions over time, capturing key points over the course of task completion (Fig. 6), conveying the changes in attentional distribution with foci growing and shrinking over time.

The proposed model thus integrates gestures as an additional modality within the focus of attention, which can provide a means of presenting the details of reinforcement and interference across foci during a learning experience, all while capturing the nuances of attentional inclination and redistribution.

Though useful as an abstract visualization, this distributed attention model must also account for CLT effects if it is to be useful for exploring cognition within the context of education. The next section will discuss specific alignment with established cognitive load effects and how the proposed model can explain these effects.

Distributed Attention and Cognitive Load Theory

The basic assumption of a limited WM capacity has led to research into the implications of WM capacity constraints for learners and learning. Specifically, CLT was developed (Sweller 1988; Sweller and Chandler 1991; Sweller et al. 1998) as a branch of educational psychology that seeks to inform educational practice through empirically supported instructional interventions that account for the WM demands (load) of teaching, learning, and instruction. Building upon Miller's limited WM capacity of "7 plus or minus 2" pieces of information, CLT recast WM terminology to describe units of information as "elements" of information that interact with each other to create schematic links and support WM processing (Halford et al. 1998; Sweller et al. 2011). For example, solving a simple math equation requires the cognitive "juggling" of multiple elements maintained and processed in WM, such as numbers and their meaning, rules governing how the numbers are to be processed, and the interpretation of symbols required to solve a presented equation. During information processing (e.g., learning), as the number of interacting elements being processed increases, cognitive load will also

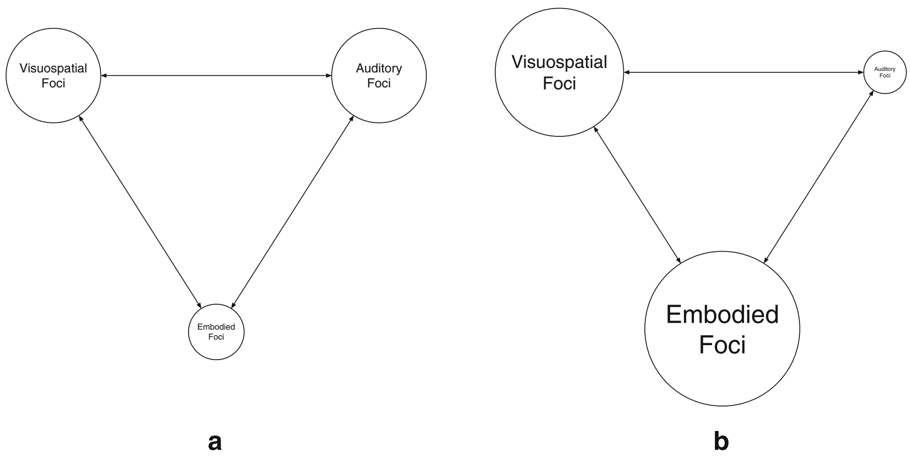


Fig. 5 **a**, **b** Examples of individual differences in attentional inclination

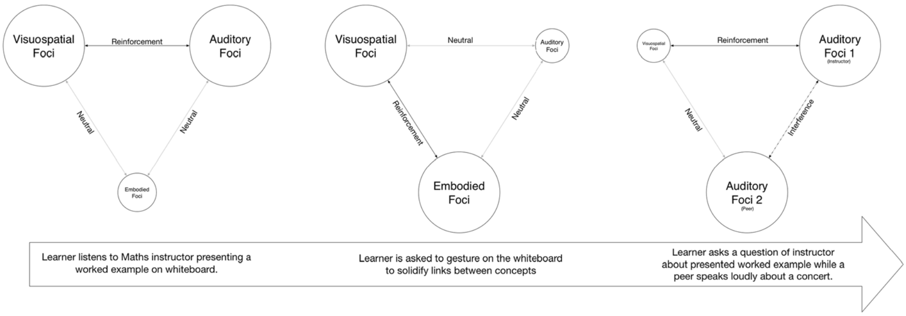


Fig. 6 Examples of key moments during learning episode

increase until WM resources reach their limit. A cognitive load at this level is not considered negative, although when WM capacity is exceeded *cognitive overload* occurs. Cognitive overload results in the loss of one or more elements from within WM and thus an inability to process the meaning of, and relationships between, active elements with those that were lost. In most learning contexts, this loss is an impediment to successful task completion and has a negative impact on learning.

The work by Engle (2002) and colleagues (Kane et al. 2001) and others (Pascual-Leone and Baillargeon 1994; Pascual-Leone and Smith 1969), which emphasizes the role that attentional control plays in allocating resources within a limited-capacity WM system, served as a basis for the proposed integrated WM model. One of CLT's foundational assumptions is that intrinsic and extraneous cognitive loads are determined by the number of multiple interacting elements (items activated) within WM. This distributed attention model aligns with this assumption, in that an increase in the number of foci processed within the focus of attention increases the attentional (WM) resources required to reconcile and integrate these foci. Importantly, this model accounts for a broad range of CLT effects, including emerging findings exploring the human motor system's role in cognition, which will be discussed later.

The first identified effect within CLT is the *goal-free effect*, which describes an increase in learning outcomes observed when a problem is phrased without an end goal described. Generally speaking, novices solving problems in subjects such as math and physics do not know the solution of the problem they are working on. As a result, learners engage in a means-ends analysis, which requires they hold in mind the goal, the known information presented in the problem, the differences between this known information and the goal, and the strategies and rules required to solve the problem. This results in an increased cognitive load due to the required simultaneous activation, maintenance, and processing of a variety of different schemas related to the problem. In a series of studies exploring the effects of this means-ends analysis problem-solving process, Sweller and colleagues (Ayres 1993; Ayres and Sweller 1990; Sweller and Levine 1982; Sweller 1988) found that when learners were provided intermediate steps to solve without an end goal—as opposed to a whole problem with a specific goal—learners were better able to find solutions through the exploration of these intermediary. Within a CLT framework, this effect is explained in terms of multiple elements interacting during the means-ends analysis process. When intermediate steps are introduced and the goal removed, element interactivity and therefore cognitive load are decreased, allowing WM resources to be freed to focus more on the steps involved in how to solve the problem (germane load). In terms of the proposed model, this effect can be explained through a reduction in the number of foci activated and the associated attentional resources required to engage in means-ends analysis within the effortful focus of attention (inner circle of the model). For example, to solve a problem, a learner would need to activate, maintain, and process multiple foci related to the known elements of the problem, the goal, the current state of their solution in relation to the goal, and the rules and strategies needed to reach that goal. When the goal is removed and intermediate steps are the only stage the learner is required to solve, limited attentional resources that were previously devoted to processing the goal, its relationship to the problem, and its current state of completion are removed. This frees limited resources to focus attention on exploring the nature of the problem itself, including processing relationships between elements and how they fit into each stage of the problem, thus fostering more robust schema creation and integration with existing expertise.

The *split-attention effect* is related to situations in which learners are exposed to instructional diagrams with mutually referring text and images separated spatially. This forces the learner to split their attention to look back and forth between the diagram and explanatory text (Chandler and Sweller 1992; Sweller and Chandler 1991). The mental integration of both information sources needed to understand the learning materials imposes a high demand on WM and can impact negatively on learning. In contrast, research shows that the associated high cognitive load can be reduced and learning better facilitated by spatially integrating the text into the diagrams (Chandler and Sweller 1992; Sweller and Chandler 1991) or by replacing the visual text with spoken text (the modality effect; Mousavi et al. 1995; Tindall-Ford et al. 1997). Mayer's (2005) cognitive theory of multimedia learning (CTML) posits a similar, yet more generally applicable concept called the "spatial contiguity principle" which asserts that when related text and imagery are displayed near to each other, learners are able to engage more deeply with the materials (for a meta-analysis of both principles, see Schroeder and Cenkci 2018). The distributed attention model explains the split-attention effect in terms of the cognitive cost of visual search between different foci. When materials are not integrated, activated schema for each foci must be effortfully maintained within the focus of attention (inner circle of the model) while the learner engages in visual search to seek out related information (Schmidt-Weigand et al. 2010), thus decreasing available attentional resources and increasing cognitive load. One study demonstrated a decrease in gaze shifts when learners were exposed to non-integrated materials (Bauhoff et al. 2012). These reduced gaze shifts suggest a longer duration of visual search, during which attentional resources must be devoted to maintaining foci related to the task. Florax and Ploetzner (2010) further suggest that learning can be facilitated through a well-organized integrated format intended to reduce visual search. Other studies which use eye tracking to explore spatial contiguity suggest that integrated materials result in increased eye movements (gaze shifts) between related text and images due to a reduced requirement for visual search. This provides learners more opportunities to establish relationships between the two foci (Holsanova et al. 2009; Johnson and Mayer 2012). By integrating text and images, the attentional resources required for schema maintenance when visual search is being undertaken are reduced along with extraneous cognitive load, allowing WM resources to better focus on establishing relationships in the support of learning (germane load).

Derived from the split-attention effect is the modality effect, which occurs in mixed mode instruction when visual learning materials (such as diagrams) are supplemented with complementary auditory information (e.g., a verbal statement in place of written text). This is more effective for learning than single-modality instruction, because the combination of modalities have been shown to make more effective use of WM resources by reinforcing each other (Mousavi et al. 1995; Tindall-Ford et al. 1997). One explanation for the modality effect suggests that learning materials presented in both written and pictorial formats induce split attention, which results in interference within the visual modality, making it challenging to integrate this information (Low and Sweller 2005). By replacing the written statements with spoken statements, split attention is reduced or negated due to the benefit of paired auditory statements with pictorial information. This benefit is often explained through the lens of separate slave systems defined in Baddeley and Hitch's multicomponent model of WM or Penney's (1989) *separate streams* hypothesis whereby the WM system is comprised of dedicated systems for processing visual and auditory information independent of one another, offloading information from one channel to both channels and allowing for concurrent processing and a reduction of cognitive load (Ginns 2005). Mayer's cognitive theory of multimedia learning (Mayer 1997, 2002, 2009; Mayer and Anderson 1992; Mayer and Moreno 1998, 2003) refers to this explanation as the *visuospatial load hypothesis* (Mayer 2001). Rummer et al. (2010) argue this explanation is not compatible with Baddeley and Hitch's model due to the phonological loop being responsible for both written and verbal (auditory) information retention and processing, thus negating the benefit. Another explanation for the modality effect which is also used to describe the split-attention effect is the *contiguity assumption* (Ginns 2006; Moreno & Mayer, 1998; Rummer et al. 2011) which refers to the temporal delay between activated information in memory that occurs after the immediate perception of separated text and pictures. In the case of the modality effect, the simultaneous exposure to auditory and visual information reduces or eliminates this delay, while more efficiently allocating WM resources for processing and learning. Further, Rummer et al. (2010) suggest a third explanation based on the *auditory recency effect* (Penney 1989), which describes when auditory information is more readily retained through sensory perception than visual information, limited in scope to the most recently heard item. It is argued that this internal verbal *echo* reduces limited WM resources required to integrate auditory and visual information as the auditory information is retained in a more salient manner.

The split-attention and modality effects continue to be explored, discussed, and debated through the lens of differing explanations for these effects rooted in perception and WM processing. The distributed attention model presented in this paper explains the modality effect by reframing established assumptions. Instead of simultaneous processing in separate subsystems dedicated to each modality, the distributed model frames these explanations through the allocation of attentional resources combined and distributed across modalities within a single, limited-capacity focus of attention system. When a learner is exposed to mutually referential auditory and visual information such as a diagram with a supporting audio statement, cognitive load is reduced in two ways. First, activated foci combined across modalities within the effortful focus of attention (inner circle of the model) do not need to be reconciled as they are not in the same modality; thus, visuospatial load hypothesis still applies in the proposed model, retaining the assumption that interference can occur between two foci of in the same modality (see Fig. 6). Further, activation of foci in a connected and distributed manner allows for more efficient processing of relationships and creation of unitary representations within WM. Second, when learners are exposed to multimodal learning materials, automatic and effortless activation of foci across modalities occurs within the periphery of attention (outer ring of the model), simultaneously, and inclusive of a more robust activation of immediate auditory information as described by the auditory recency effect. As there is little to no delay between activations of these foci, attentional resources that would be devoted to maintaining activated foci and searching for other related foci within the same modality are no longer required; thus, the contiguity assumption still holds. When considering these differing explanations for the modality effect, the distributed attention model is able to account for each through the lens of attentional focus and distribution, which succinctly describes the activation, combination, and integration of attentional foci across modalities, resulting in a more efficient allocation of attentional resources, which supports learning and schema creation.

As a further extension of the split-attention effect, the redundancy effect (Chandler & Sweller, 1991; Kalyuga et al. 1999) shows that when information is duplicated–either in the same modality or in a different one–learning can be negatively impacted. The redundancy effect and *expertise reversal effects* refer to a nullification of cognitive efficiencies through the presentation of redundant information or existing expertise, respectively (Kalyuga et al. 2003; Kalyuga et al. 1999). A study by Kalyuga et al. (1999) demonstrated that when learners were presented with multimedia instructions related to mechanical engineering, learners who were presented with auditory statements corresponding to visual diagrams outperformed those who were given just written statements or combined written and auditory statements. According to Baddeley and Hitch's (1974) model, these compatible sources of information would be activated and processed in separate slave systems, but this model does not appear to account for why an impairment in learning should occur in this manner. Alternatively, Cowan's (1988, 1995) theorizing of a single WM system can explain this redundancy in terms of two or more sources of duplicated information requiring reconciliation and integration within the scope and control of attention, all of which accumulates to tax available WM resources. In terms of the distributed attention model, the explanation is similar to Cowan's in that the need to reconcile redundant information across two different foci wastes WM resources in the pursuit of reconciling what has already been presented. With regard to expertise reversal, if students are presented with information they have already mastered, existing schemas related to this expertise are automatically retrieved from LTM and activated within the effortful focus of attention (inner circle of the model). This results in an increase in attentional resources devoted to reconciling existing knowledge with unnecessarily presented foci (extraneous load) and may reduce learning outcomes.

For each described CLT effect, interpretation does not require allusion to a multidomain WM system or the separation of storage from processing. Instead, the processing of information (any information) places similar demand on attentional resources for working with information in WM as processing of that information. As a final example, consider simple multiplication; before this process is automated, mentally calculating the product of two digits is more cognitively demanding than simply holding those digits in mind. According to this model, this is because the number and complexity of elements (including the complexity of mental manipulations to be performed on these elements), and their interactivity, all draw upon a unitary but limited attentional resource. Differences in performance across domains are explained by our different levels of expertise within those domains, just as expertise playing an instrument does not automatically transfer to expertise with another. In these propositions, CLT effects can be explained both comprehensively and simply, providing a model that does not need to be revised with each new domain or effect uncovered by CLT researchers.

In this section, we have discussed how the proposed distributed attention model may explain well-established effects and instructional principles identified over many years of CLT research. While the models of Baddeley and Hitch (1974), Cowan (1988), and Mayer (2005) are able to explain specific CLT effects in terms of perception, processing, and storage where appropriate, these models are individually unable to account for every CLT effect in a flexible and robust manner. This issue is further exacerbated by recent CLT findings regarding the human motor system, which these WM models have not yet explicitly attempted to reconcile. In the following sections, an overview of research into the human motor system and its effect on learning is discussed, including how the proposed model integrates these findings in contrast to established models, while also positioning gestures and human movements as an additional modality.

How the Human Motor System Supports Learning and Cognition

When considering the reinforcement or interference between sources of information across different modalities, educational and psychological research has historically focused on isolating auditory and visual information through passive exposure to these sources. However, a question remains about whether and how other modalities may also play a role in learning and cognition. For instance, most learners are able to involve their own motor system in the learning process, by using either fine motor movements, such as hand and finger gestures (e.g., finger counting, tracing, and pointing), or gross motor movements, such as whole body movements (e.g., mimicking others and enacting concepts). It is thus plausible that human movement may constitute one of many additional modalities that have yet to be comprehensively considered in relation to existing WM models.

A *human movement effect* was recently identified within CLT, suggesting that learning procedural motor tasks from dynamic visualizations, such as animations, can be more effective than learning from static visualizations when the transient animated information incorporates human movement (Ayres et al. 2009; Castro-Alonso et al. 2014; Paas and Sweller 2012; Wong et al. 2009; see also Höffler and Leutner 2007). For instance, Wong et al. (2009) found that when learners observed animations of paper folding from a first-person perspective, they were able to replicate the folding task with increased proficiency when compared with those who viewed static images of each step. While the animations did not show hands, they presented paper folding as if hands were present, suggesting a potential role for gestures in WM processes.

The field of embodied cognition (Foglia and Wilson 2013) is a separate area of research that is viewed as a meaningful theoretical companion to CLT (Sweller et al. 2011). Embodied cognition asserts that cognitive processes, including information processing and learning, are inextricably linked with all forms of sensory input (not just sight and sound), including physical and environmental experiences of an individual. In a basic sense, embodied cognition frames all cognition as being intrinsically linked with sensory and motor functions within the environment, including gestures and other human movements (Barsalou 1999). Although not originally rooted in educational contexts, embodied cognition presents a meaningful anchor for existing research in CLT focusing on human movement and gestures, allowing a theoretical link to be made (Choi et al. 2014; Paas and Sweller 2012). Framed in this way, a learner's motor functions, including gestures and tactile experiences, play a similar role to (and introduce similar effects as) visual and auditory information in learning. In one study of note, tangible physical user interfaces in a multimedia learning context were investigated for learning anatomy, with results demonstrating increased learning outcomes when compared to a traditional mouse and computer display experience (Skulmowski et al. 2016). This, along with studies on the use of cognitive load measurement for embodied learning (for a review, see Skulmowski and Rey 2017), has begun to position physical experiences as an established area of inquiry along with cognitive load.

Within a CLT framework, the use of iconic hand gestures (that is, those that meaningfully represent objects, visuospatial traits, or actions) has been shown to benefit learning of a foreign language (Macedonia and Klimesch 2014; Mavilidi et al. 2015). When preschool-aged children produced iconic gestures representing an action that matches its foreign language word/phrase, such as acting out the word for "swim" while learning the word in Italian (*nuotare*), children's learning outcomes were enhanced and when full-body movement was compared with arm and hand gesturing while sitting, the learners engaged in full-body movement benefitted further (Mavilidi et al. 2015). These same types of gestures also seem capable of supporting mimicry. When observing an adult demonstrates how a toy worked through an iconic gesture (pretending to act on an object nearby without touching it), an infant's ability to successfully operate the toy increased (Novack et al. 2015). These gesturing effects are also supported by findings in neurology around the mirror neuron system (MNS). The MNS is a neurological system located in the premotor cortex that activates when humans and other nonhuman primates observe an action performed by another individual. This observation is thought to cognitively prime the observer to perform the same action (for a review, see Rizzolatti and Craighero 2004), a finding that serves as a partial explanation for the human movement effect (Van Gog et al. 2009).

In contrast to research in more-controlled environments such as research labs, a body of literature has grown around exploring the use of gestures and their effects in more traditional learning environments, such as classrooms and preschools. Research by Goldin-Meadow and others (Cook et al. 2013; Cook et al. 2008; Goldin-Meadow 2009; Goldin-Meadow et al. 2009; Novack and Goldin-Meadow 2015), exploring the role of pointing, iconic gesture, and metaphoric gestures (i.e., gestures that represent abstract ideas) in math and language learning, has demonstrated that these gestures provide a support mechanism that improves learning outcomes. In explaining these findings, it has been suggested that gestures could be used for the "cognitive offloading" of information during problem solving, leading to a reduction in cognitive load (Cook et al. 2012; Goldin-Meadow et al. 2001; Ping and Goldin-Meadow 2010; Risko and Gilbert 2016). These findings provide further evidence for the positive role that gestures may play in classroom-based learning, positioning gestures as an additional contributor to learning alongside more traditional auditory and visual inputs. How gesturing effects are explained by existing and emerging theories of WM, however, requires a discussion of gestures from a psychological perspective.

Efforts have begun to explain these motor-related findings through the construct of WM. Research exploring how human movement relates to WM (Engelkamp 1995; Engelkamp et al. 2005; Engelkamp et al. 1994) has focused specifically on the memory traces of action performance. In one study, participants were exposed to both ordinary and "bizarre" action phrases (actions that are novel or surreal, such as "plant the hammer"), with half of participants learning the phrases with only verbal prompts and the other half performing a motor task aligned with the phrase. When participants were given a list of phrases and asked to identify which of them they had learned, those who performed the actions during the learning phase demonstrated increased recognition (Engelkamp et al. 1994). While Engelkamp explained these findings through an extension of dual coding theory (DCT), it is nevertheless informative for research focusing on WM. DCT posits that mental representations can be visual and verbal in nature, processed through two distinct and co-reinforcing channels (Paivio and Okovita 1971; see also Clark and Paivio 1991). However, Wilson (2001) observes that although memory coding for human movement is empirically supported, "recent models have shown a consistent trend away from sensorimotor representations" (p. 44). This raises questions about the appropriateness of CLT referencing WM models of Baddeley and Hitch (1974) and Cowan (1988) to explain the effects of human movement on learning. As such, in the next section, Baddeley and Hitch's WM model will be discussed in terms of whether gestures and movement might be reconcilable with the multicomponent model.

Gestures as an Additional Modality

In contrast to Baddeley and Hitch's focus on verbal and visual-spatial information, there is a growing body of evidence which suggests that motor information may constitute an additional modality that can also occupy WM's limited resources. For instance, Wilson and Emmorey's (1997) work with speakers of American Sign Language (ASL) found that when participants' rehearsal of signs included interference (performing a nonsense sign), their recall of actual signed concepts was reduced. Paralleling CLT findings of a negative impact of extraneous information, the nonsense signs reduced overall performance (that is, they placed demands on WM, leading to cognitive overload and loss of information). Based on these results, the researchers noted, "the working memory system of a deaf ASL signer contains a rehearsal loop that possesses many of the structural properties of the phonological loop for speech" (Wilson and Emmorey 1997, p. 319). It is thus important to consider that while the phonological loop is traditionally tied to spoken language, this is not the only form of communication available. In a review by Rudner et al. (2009), the authors concluded that sign language and spoken language, from both a WM and neurological perspective, are similar. However, it was found that deaf individuals who signed from a young age exhibit an inclination towards spatial organization of information, whereas hearing participants often prefer temporal organization (Cumming and Rodda 1985). For example, spoken languages such as English are linear in nature with a sequenced ordering of sounds to represent ideas, whereas signed languages such as ASL rely on the spatial referents and semantic structures of sequenced gestures, body movements, and facial expressions to convey meaning. This means that experience in a particular language can play a role in WM processing through an inclination towards the modality of that language (spoken or signed). Wilson and Emmorey (2003) also found that when hearing and deaf participants were asked to recall concepts presented in their first language, only the deaf participants were sensitive to interference inputs presented in ASL. This mirrors similar findings for hearing participants presented with written interference tasks (e.g., reading while engaged in a word-span task; Turner and Engle 1989).

These findings suggest that signed languages use similar cognitive processes to spoken and written languages and are similarly susceptible to interference and reinforcement. Further, proficiency in a particular language influences WM processing in favor of that language's modality (e.g., visuospatial, auditory, gesture), which, by extension, can influence the application and allocation of WM resources. These findings are important for conceptions of separate WM systems for different modalities: first, because they contribute to previous work in psychology through the investigation of modality isolation and cognitive interference, and second, because it indicates that gestures and human movement may be considered both visuospatial and *verbal*, so relegating them to the phonological loop or visuospatial sketchpad alone may not be viable.

Given that signed languages and spoken languages can share similar WM processes, the distinction between Baddeley and Hitch's phonological loop and visuospatial sketchpad in terms of where gestures fit is unclear. Baddeley (2012) himself even raised questions about how other physical experiences such as tactile and nonspeech kinesthetic experiences could be integrated within his existing model, and suggested how slave systems could possibly exist for even more senses and even more types of stimuli input (see Fig. 7). However, how these might all interact, and how this would account for reinforcement and interference effects across different combinations of modalities, remains unclear.

Instead of investigating and revising an ever-expanding list of WM systems, their functions, and their interactions, it is contended that distributed attention model presented in this paper can better account for the inclusion of multiple sensory and processing domains, through a single and integrated attentional resource that is involved in the activation of visual, phonological, and embodied information (within the scope of attention), while also accounting for individual differences in experience and expertise (through the control of attention).

The distributed attention model thus provides a framework for integrating gestures and human movement as a modality which, much like auditory and visual information, provide learning gains through reinforcement or learning deficits through the interference between multiple sources of information within the focus of attention. The human movement effect is therefore framed as the freeing of WM resources for the creation of schema when learning is supported by movement (reinforcement).

In this section, recent findings in embodied cognition and the human movement effect have been presented, including how the human motor system can support learning. A discussion of how these findings may be integrated into established models of WM presents a unique challenge while also speaking to the affordances of the proposed model for the integration of gestures and human movement as an additional modality. In the next section, limitations of the proposed model are discussed including its scope, potential for future application, and opportunities for validation.

Limitations of the Proposed Model

Studies investigating cognitive processes across multiple sensory modalities such as auditory and visuospatial have traditionally explored these areas from the perspective of cognitive interference and reinforcement. Isolating specific modalities (i.e., attentional foci) has provided

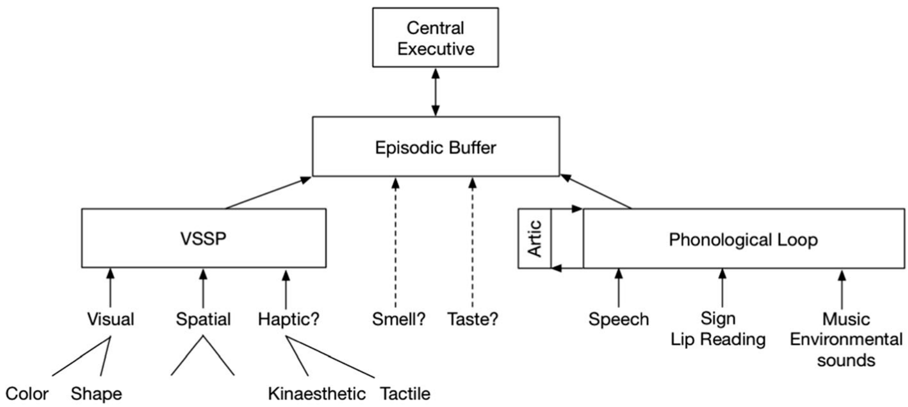


Fig. 7 Diagram based on Baddeley's speculative view of the flow of information from perception to working memory (Baddeley 2012)

many theoretical and practical advances both in psychology and learning sciences. The primary limitation of the distributed attention model presented in this paper is that it is not formulated to explore modalities in isolation, but instead, the interaction between them. While it may not be possible to explore isolated modalities through the lens of this model, it does afford the opportunity to investigate a learners' ability to control the focus and distribution of attention across modalities. The dynamic nature of this model serves primarily as a theoretical framework for representing individual differences and changes in WM resource allocation over time, which itself may be presently challenging to quantify. In addition, the proposed model does not explicitly account for initial perceptual limitations but focuses more on post-stimulus processing within WM. The intentional dynamic nature of the model could, however, be used to indicate initial attentional inclination and attentional focus immediately post-stimulus to explore the relationships between modalities in future research.

To investigate further, future studies may choose to build upon measures of attentional focus, including eye tracking with an emphasis on tagging or categorizing objects of visual focus over time. Recent advances in motion tracking technology may also allow for capturing of gross full body movements, as well as fine hand and finger movements. As the ability to capture information related to the human motor system advances, attentional focus on auditory information may still prove challenging to quantify in educational settings. Collecting subjective learner reflections on external or internal differences in attentional focus, including distractions, cognitive or emotional challenges, and previous cultural experiences may also provide a more nuanced picture of cognitive processes during task completion. It may also prove beneficial to investigate distribution of attention through a combination of these measures along with subjective learner perception of their own inclination towards specific attentional foci.

Conclusion

Given uncertainty around how recent gesture effects could be reconciled within the WM frameworks that CLT researchers normally defer to, we attempted a reconciliation and also cast our attention more widely. In doing so, an integrated distributed attention model of WM was proposed, which accounts for the breadth of CLT findings in a comprehensive but parsimonious way. This model integrates WM principles and insights from theorists including Cowan, Engle, Pascual-Leone, and others. By presenting a unitary attentional (WM) resource that can integrate information from multiple sources and modalities and can adjust the distribution of attentional focus during processing, this model may assist in explaining, clarifying, recasting, and supporting CLT findings now and in the future.

References

Agostinho, S., Tindall-Ford, S., Ginns, P., Howard, S. J., Leahy, W., & Paas, F. (2015). Giving learning a helping hand: finger tracing of temperature graphs on an iPad. *Educational Psychology Review, 27*(3), 427–443. https://doi.org/10.1007/s10648-015-9315-5.

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* (Vol. 2, pp. 89–195). London: Academic.

Ayres, P. L. (1993). Why goal-free problems can facilitate learning. *Contemporary Educational Psychology, 18*(3), 376–381. https://doi.org/10.1006/ceps.1993.1027.

Ayres, P., Marcus, N., Chan, C., & Qian, N. (2009). Learning hand manipulative tasks: when instructional animations are superior to equivalent static representations. *Computers in Human Behavior, 25*(2), 348–353. https://doi.org/10.1016/j.chb.2008.12.013.

Ayres, P., & Sweller, J. (1990). Locus of difficulty in multi-stage mathematics problems (Ph.D. thesis). *The American Journal of Psychology, 103*(2), 167–193.

Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society B: Biological Sciences, 302*(1110), 311–324. https://doi.org/10.1098/rstb.1983.0057.

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

Baddeley, A. D. (1992). Working memory. *Science, 255*(5044), 556–559. https://doi.org/10.1126/science.1736359.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2.

Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829–839. https://doi.org/10.1038/nrn1201.

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology, 63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic.

Baddeley, A. D., & Lieberman, K. (1980). Spatial working memory. In R. S. Nickerson (Ed.), *Attention and performance VIII* (pp. 521–539). Hillsdale: Lawrence Erlbaum Associates, Inc.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences, 22*(04), 637–660.

Bauhoff, V., Huff, M., & Schwan, S. (2012). Distance matters: spatial contiguity effects as trade-off between gaze switches and memory load. *Applied Cognitive Psychology, 26*(6), 863–871. https://doi.org/10.1002/acp.2887.

Broadbent, D. E. (1958). Perception and Communication. New York: Pergamon.

Broadbent, D. E. (1982). Task combination and selective intake of information. *Acta Psychologica, 50*(3), 253–290.

Brucker, B., Ehlis, A.-C., Häußinger, F. B., Fallgatter, A. J., & Gerjets, P. (2015). Watching corresponding gestures facilitates learning with animations by activating human mirror-neurons. *An fNIRS study, 36*(C), 27–37. https://doi.org/10.1016/j.learninstruc.2014.11.003.

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2014). Learning from observing hands in static and animated versions of non-manipulative tasks. *Learning and Instruction, 34*, 11–22.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*(2), 233–246. https://doi.org/10.1111/j.2044-8279.1992.tb01017.x.

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2017). Extending cognitive load theory to incorporate working memory resource depletion: evidence from the spacing effect. *Educational Psychology Review, 61*(2), 1–19. https://doi.org/10.1007/s10648-017-9426-2.

Choi, H. H., Van Merrienboer, J. J. G., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: towards a new model of cognitive load. *Educational Psychology Review, 26*(2), 225–244.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137*(4), 706–723. https://doi.org/10.1037/a0013157.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2013). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *Journal of Experimental Psychology: General, 143*, 694–709. https://doi.org/10.1037/a0033861.supp.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149–210.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769–786. https://doi.org/10.3758/BF03196772.

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development, 84*(6), 1863–1871. https://doi.org/10.1111/cdev.12097.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*(2), 1047–1058. https://doi.org/10.1016/j.cognition.2007.04.010.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes, 27*(4), 594–610. https://doi.org/10.1080/01690965.2011.567074.
Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163–191.
Cowan, N. (1995). *Attention and memory: an integrated framework*. New York: Oxford University Press; Oxford: Clarendon, 1995.
Cowan, N. (2000). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–185.
Cowan, N. (2010). The magical mystery four. *Current Directions in Psychological Science, 19*(1), 51–57. https://doi.org/10.1177/0963721409359277.
Cumming, C. E., & Rodda, M. (1985). The effects of auditory deprivation on successive processing. *Canadian Journal of Behavioural Science / Revue Canadienne des Sciences du Comportement, 17*(3), 232–245.
Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: a deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 53, pp. 63–147). New York: Academic. https://doi.org/10.1016/S0079-7421(10)53003-2.
De Groot, A. (1965). Thought and choice in chess. The Hague: Mouton. Original work published 1946.
De Koning, B., & Tabbers, H. K. (2013). Gestures in instructional animations: a helping hand to understanding non-human movements? *Applied Cognitive Psychology, 27*, 683–689.
Ebbinghaus, H. (1885/1964). Memory: a contribution to experimental psychology. Oxford: Dover.
Engelkamp, J. (1995). Visual imagery and enactment of actions in memory. *British Journal of Psychology, 86*(2), 227–240. https://doi.org/10.1111/j.2044-8295.1995.tb02558.x.
Engelkamp, J., Seiler, K. H., & Zimmer, H. D. (2005). Differential relational encoding of categorical information in memory for action events. *Memory & Cognition, 33*(3), 371–389.
Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: self-performing during recognition. *Memory & Cognition, 22*(1), 34–39. https://doi.org/10.3758/BF03202759.
Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160.
Florax, M., & Ploetzner, R. (2010). What contributes to the split-attention effect? The role of text segmentation, picture labelling, and spatial proximity. *Learning and Instruction, 20*(3), 216–224. https://doi.org/10.1016/j.learninstruc.2009.02.021.
Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(3), 319–325.
Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*(4), 313–331. https://doi.org/10.1016/j.learninstruc.2005.07.001.
Ginns, P. (2006). Integrating information: a meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*(6), 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001.
Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology, 28*(2), 266–273. https://doi.org/10.1002/acp.2997.
Goldin-Meadow, S. (2009). How gesture promotes learning throughout childhood. *Child Development Perspectives, 3*(2), 106–111.
Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. (2009). Gesturing gives children new ideas about math. *Psychological Science, 20*(3), 267–272. https://doi.org/10.1086/659964?ref=search-gateway:4d5c7866182d0f01d9bef79f563ad121.
Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychological Science, 12*(6), 516–522.
Healey, M. K., Hasher, L., & Danilova, E. (2011). The stability of working memory: do previous tasks influence complex span? *Journal of Experimental Psychology: General, 140,* 573–585. https://doi.org/10.1037/a0024587.
Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: individual differences in visual attention control. *Journal of Experimental Psychology: General, 136*(2), 217–240. https://doi.org/10.1037/0096-3445.136.2.217.
Hu, F. T., Ginns, P., & Bobis, J. (2014). Does tracing worked examples enhance geometry learning? *Australian Journal of Educational Developmental Psychology, 14*, 45–49.
Kalyuga, S., Ayres, P., Chandler, P. A., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4.
Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 25*(S1), 123–S144. https://doi.org/10.1002/acp.1773.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169–183. https://doi.org/10.1037//0096-3445.130.2.169.

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: long-term benefits for factual and higher-level learning. *Learning and Instruction, 36*, 38–45. https://doi.org/10.1016/j.learninstruc.2014.11.001.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*, 803–831.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: a meta-analysis. *Learning an Instruction, 17*(6), 722–738.

Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: the role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology, 23*(9), 1215–1226. https://doi.org/10.1002/acp.1525.

Juan Pascual-Leone, (1970) A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 32* 301–345

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18*(2), 178–191. https://doi.org/10.1037/a0026923.

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147–158). New York: Cambridge University Press.

Macedonia, M., & Klimesch, W. (2014). Long-term effects of gestures on memory for foreign language words trained in the classroom. *Mind, Brain, and Education, 8*(2), 74–88.

Mavilidi, M. F., Okely, A. D., Chandler, P. A., Cliff, D. P., & Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educational Psychology Review, 27*(3), 413–426. https://doi.org/10.1007/s10648-015-9337-z.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1–19.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation, 41*, 85–139.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning* (p. 2009). Cambridge: Cambridge University Press.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology, 84*(4), 444–452. https://doi.org/10.1037/0022-0663.84.4.444.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320. https://doi.org/10.1037/0022-0663.90.2.312.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. https://doi.org/10.1037/h0043158.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319–334. https://doi.org/10.1037/0022-0663.87.2.319.

Novack, M., & Goldin-Meadow, S. (2015). Learning from gesture: how our hands change our minds. *Educational Psychology Review, 27*(3), 405–412.

Novack, M., Goldin-Meadow, S., & Woodward, A. L. (2015). Learning from gesture: how early does it happen? *Cognition, 142*, 138–147.

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*(1), 27–45.

Paivio, A., & Okovita, H. W. (1971). Word imagery modalities and associative learning in blind and sighted subjects. *Journal of Verbal Learning and Verbal Behavior, 10*(5), 506–510.

Pascual-Leone, J., & Baillargeon, R. (1994). Developmental measurement of mental attention. *International Journal of Behavioral Development, 17*(1), 161–200. https://doi.org/10.1177/016502549401700110.

Pascual-Leone, J., & Smith, J. (1969). The encoding and decoding of symbols by children: a new experimental paradigm and a neo-Piagetian model. *Journal of Experimental Child Psychology, 8*(2), 328–355. https://doi.org/10.1016/0022-0965(69)90107-6.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition, 17*(4), 398–422. https://doi.org/10.3758/BF03202613.

Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science, 34*(4), 602–619.

Pouw, W., Mavilidi, M.-F., Van Gog, T., & Paas, F. (2016). Gesturing during mental problem solving reduces eye movements, especially for individuals with lower visual working memory capacity. *Cognitive Processing, 17*(3), 269–277. https://doi.org/10.1007/s10339-016-0757-6.

Pouw, W. T. J. L., de Nooijer, J. A., Van Gog, T., Zwaan, R. A. & Paas, F. (2014a). Toward a more embedded/extended perspective on the cognitive function of gestures. *Frontiers in Psychology, 5*, 1–14. https://doi.org/10.3389/fpsyg.2014.00359.

Pouw, W., Van Gog, T., & Paas, F. (2014b). An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review, 26*(1), 51–72.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230.

Rhodes, G. (1987). Auditory attention and the representation of spatial information. *Perception & Psychophysics, 42*(1), 1–14.

Rudner, M., Andin, J., & Rönnberg, J. (2009). Working memory, deafness and sign language. *Scandinavian Journal of Psychology, 50*(5), 495–505. https://doi.org/10.1111/j.1467-9450.2009.00744.x.

Rummer, R., Schweppe, J., Fürstenberg, A., Seufert, T., & Brünken, R. (2010). Working memory interference during processing texts and pictures: implications for the explanation of the modality effect. *Applied Cognitive Psychology, 24*(2), 164–176. https://doi.org/10.1002/acp.1546.

Rummer, R., Schweppe, J., Fürstenberg, A., Scheiter, K., & Zindler, A. (2011). The perceptual basis of the modality effect in multimedia learning. *Journal of Experimental Psychology: Applied, 17*(2), 159–173. https://doi.org/10.1037/a0023588.

Schmalenbach, S. B., Billino, J., Kircher, T., van Kemenade, B. M., & Straube, B. (2017). Links between gestures and multisensory processing: individual differences suggest a compensation mechanism. *Frontiers in Psychology, 8*, 635–638. https://doi.org/10.3389/fpsyg.2017.01828.

Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General, 136*(2), 241–255. https://doi.org/10.1037/0096-3445.136.2.241.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). Explaining the modality and contiguity effects: new insights from investigating students' viewing behaviour. *Applied Cognitive Psychology, 24*(2), 226–237. https://doi.org/10.1002/acp.1554.

Schroeder, N. L., & Cenkci, A. T. (2018). Spatial contiguity and spatial split-attention effects in multimedia learning environments: a meta-analysis. *Educational Psychology Review, 30*(3), 679–701.

Skulmowski, A., Pradel, S., Kühnert, T., Brunnett, G., & Rey, G. D. (2016). Embodied learning using a tangible user interface: the effects of haptic perception and selective pointing on a spatial learning task. *Computers & Education, 92-93*(c), 64–75. https://doi.org/10.1016/j.compedu.2015.10.011.

Skulmowski, A., & Rey, G. D. (2017). Measuring cognitive load in embodied learning settings. *Frontiers in Psychology, 8*, 1–6. https://doi.org/10.3389/fpsyg.2017.01191.

Smith, D., Davis, B., Niu, K., Healy, E., Bonilha, L., Fridriksson, J., Morgan, P. S., & Rorden, C. (2009). Spatial attention evokes similar activation patterns for visual and auditory stimuli. *Journal of Cognitive Neuroscience, 22*, 347–361.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4.

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

Sweller, J., & Chandler, P. A. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332. https://doi.org/10.1207/s1532690xci0804_2.

Sweller, J., Levine, M., 1982. (1982). Effects of goal specificity on means–ends analysis and learning. Journal of Experimental Psychology, 8(5), 463–474.

Sweller, J., Van Merrienboer, J. J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

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

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127–154.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Bulletin, 114*(1), 104–132. https://doi.org/10.1037/0033-295X.114.1.104.

Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror neuron system and observational learning: implications for the effectiveness of dynamic visualizations. *Educational Psychology Review, 21*, 21–30.
Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review, 8*(1), 44–57.
Wilson, M., & Emmorey, K. (1997). A visuospatial "phonological loop" in working memory: evidence from American Sign Language. *Memory & Cognition, 25*(3), 313–320. https://doi.org/10.3758/BF03211287.
Wilson, M., & Emmorey, K. (2003). The effect of irrelevant visual input on working memory for sign language. *Journal of Deaf Studies and Deaf Education, 8*(2), 97–103.
Wong, A., Marcus, N., Ayres, P., Smith, L., Cooper, G. A., Paas, F., & Sweller, J. (2009). Instructional animations can be superior to statics when learning human motor skills. *Computers in Human Behavior, 25*(2), 339–347. https://doi.org/10.1016/j.chb.2008.12.012.

Affiliations

Stoo Sepp[1] • Steven J. Howard[1] • Sharon Tindall-Ford[1] • Shirley Agostinho[1] • Fred Paas[1,2]

1 Early Start and School of Education, University of Wollongong, Wollongong, NSW 2522, Australia

2 Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands