Working Memory Underpins Cognitive Development, Learning, and Education

Working Memory Underpins Cognitive Development, Learning, and Education

Nelson Cowan
University of Missouri

Abstract

Working memory is the retention of a small amount of information in a readily accessible form. It facilitates planning, comprehension, reasoning, and problem-solving. I examine the historical roots and conceptual development of the concept and the theoretical and practical implications of current debates about working memory mechanisms. Then I explore the nature of cognitive developmental improvements in working memory, the role of working memory in learning, and some potential implications of working memory and its development for the education of children and adults. The use of working memory is quite ubiquitous in human thought, but the best way to improve education using what we know about working memory is still controversial. I hope to provide some directions for research and educational practice.

What is Working Memory? An Introduction and Review

Working memory is the small amount of information that can be held in mind and used in the execution of cognitive tasks, in contrast with long-term memory, the vast amount of information saved in one's life. Working memory is one of the most widely-used terms in psychology. It has often been connected or related to intelligence, information processing, executive function, comprehension, problem-solving, and learning, in people ranging from infancy to old age and in all sorts of animals. This concept is so omnipresent in the field that it requires careful examination both historically and in terms of definition, to establish its key characteristics and boundaries. By weaving together history, a little philosophy, and empirical work in psychology, in this opening section I hope to paint a clear picture of the concept of working memory. In subsequent sections, implications of working memory for cognitive development, learning, and education will be discussed in turn, though for these broad areas it is only feasible to touch on certain examples.

Some researchers emphasize the possibility of training working memory to improve learning and education. In this chapter, I take the complementary view that we must learn how to adjust the materials to facilitate learning and education with the working memory abilities that the learner has. Organizing knowledge, for example, reduces one's memory load because the parts don't have to be held in mind independently.

Correspondence: Nelson Cowan, Department of Psychological Sciences, University of Missouri, McAlester Hall, Columbia, MO 65211 USA, Tel. 1-573-882-4232; Fax 1-573-882-7710, CowanN@missouri.edu.

Take, for example, the possibility of doing some scouting ahead so that you will know what this article is about, making your task of reading easier. If you tried to read through the headings of this article, you might have trouble remembering them (placing them all in working memory) so as to anticipate how they fit together. If you read Figure 1, though, it is an attempt to help you organize the information. If it helps you associate the ideas to one another to build a coherent framework, it should help you read by reducing the working-memory load you experience while reading. In doing so, you are building a rich structure to associate the headings with one another in long-term memory (e.g., Ericsson & Kintsch, 1995), which reduces the number of ideas that would have to be held independently in working memory in order to remember the organization.

Early History of Working Memory Research

In 1690, John Locke distinguished between contemplation, or holding an idea in mind, and memory, or the power to revive an idea after it has disappeared from the mind (Logie, 1996). The holding in mind is limited to a few concepts at once and reflects what is now called working memory, as opposed to the possibly unlimited store of knowledge from a lifetime that is now called long-term memory. Working memory can be defined as the small amount of information that can be held in an especially accessible state and used in cognitive tasks.

Philosophers have long been interested in the limits of what can be contemplated, as noted by a leading British economist and logician, William Stanley Jevons. In an article in *Science* in 1871, he mused (p. 281): "It is well known that the mind is unable through the eye to estimate any large number of objects without counting them successively. A small number, for instance three or four, it can certainly comprehend and count by an instantaneous and apparently single act of mental attention." Then he devised a little experiment to test this hypothesis, on himself. On each trial, he casually reached into a jar full of beans, threw several beans onto a table, and tried to estimate their number without counting. After 1,027 trials, he made no errors for sets of 3 or 4 beans, with some small errors for sets of 5 beans, and with increasing magnitudes of error as a function of set size thereafter, up to 15 beans. Despite the problematic nature of the method (in that the bean thrower was also the bean judge), the finding that normal adults typically can keep in mind only about 3 or 4 items has been replicated many times in modern research, using methods similar to Jevons (e.g., Mandler & Shebo, 1982) and using many other methods (Cowan, 2001). The limited amount that could be held in mind at once played an important role in early experimental psychology, e.g., in the early experimental work of Hermann Ebbinghaus (1885/1913) and Wilhelm Wundt (1894/1998). On the American front, William James (1890) wrote about a distinction between primary memory, the items in consciousness and the trailing edge of what is perceived in the world, and secondary memory, the items in storage but not currently in consciousness. Recent investigators have considered multiple possible reasons why primary memory might be limited to just a few items at once, including biological accounts based on the need to avoid confusion between concurrent objects in memory, and evolutionary and teleological accounts based on ideas about what capacity might be ideal for learning and memory retrieval (Cowan, 2010; Sweller, 2011), but as yet the reason is unknown.

Ubiquity of the Working Memory Concept

When we say that working memory holds a small amount of *information*, by this term we may be referring to something as abstract as ideas that can be contemplated, or something as concrete as objects that can be counted (e.g., beans). The main point of information is that it is the choice of some things out of a greater set of possible things. One of the exciting aspects of working memory is that it may be important on so many different levels, and in so many different situations. When you are listening to language, you need to retain information about the beginning of the sentence until you can make sense of it. If you hear *Jean would like to visit the third building on the left* you need to recall that the actor in the sentence is Jean. Then you need to retain the verb until you know what it is she would like to visit, and you need to retain the adjective "third" until you know, third what; and all of the pieces must be put together in the right way. Without sufficient working memory, the information would be lost before you could combine it into a coherent, complete thought. As another example of how working memory is used, when doing simple arithmetic in your head, if you want to add 24 and 18 you may need to find that 4+8=12, retain the 2 while carrying the 1 over to the tens column to make 2+1+1=4 in the tens column, and integrate with the ones columns to arrive at the answer, 42. As a third example, if you are searching for your car in a parking lot, you have to remember the layout of the cars in the region you just searched so that you can avoid wasting time searching the same region again. In the jungle, a predator that turns its vision away from a scene and revisits it moments later may use working memory to detect that something in the scene has shifted; this change detection may indicate the presence of prey.

So the information in working memory can range from spoken words and printed digits to cars and future meals. It can even encompass abstract ideas. Consider whether a young child can get a good understanding of what is or is not a tiger (a matter of word category concepts, e.g., Nelson, 1974; Saltz, Soller, & Sigel, 1972). The concept is, in lay terms, a big cat with stripes. It excludes lions, which have no stripes, and it excludes zebras, which are not big cats. The child must be able to keep in mind the notion of a cat and the notion of stripes at the same time in order to grasp the tiger concept correctly. If the child thinks only of the stripes, he or she may incorrectly label a zebra as a tiger. The concept presumably starts out in working memory and, once it is learned, is transferred to long-term memory. At first, an incomplete concept might be stored in long-term memory, leading to misconceptions that are corrected later when discrepancies with further input are noticed and working memory is used to amend the concept in long-term memory. On a more abstract plane, there are more semantic issues mastered somewhat later in childhood (e.g., Clark & Garnica, 1974). The concept of *bringing* something seems to require several conditions: the person doing the bringing must have something at a location other than the speaker's location (or future planned location), and must accompany that thing to the speaker's location. You can ask the person to *bring* a salad to your house, but probably not to *take* a salad to your house (unless you are not there), and not to *send* a salad to your house (unless they are not coming along). These conditions can tax working memory. Again, the child's initial concept transferred from working memory to long-term memory may be incomplete, and amended later when discrepancies with further input are noticed.

Working Memory: The Past 64 Years

There are several modern beginnings for the working memory concept. Hebb (1949) had an outlook on temporary memory that was more neurologically based than the earlier concept of primary memory of James (1890). He spoke of ideas as mediated by assemblies of cells firing in a specific pattern for each idea or concept, and only a few cell assemblies would be active, with current neural firing, at any moment. This vision has played an important role in the field. An issue that is raised by this work is whether working memory should be identified with all of the active information that can be used in immediate memory tests, whether conscious or not, or whether it should be reserved to describe only the conscious information, more in the flavor of James. Given that working memory is a term usually used to explain behavioral outcomes rather than subjective reports, it is typically not restricted to conscious primary memory (e.g., see Baddeley, 1986; Baddeley & Hitch, 1974; Cowan, 1988). Cowan explicitly suggested that there are two aspects of working memory storage: (1) the activated portion of long-term memory, perhaps corresponding to Hebb's active cell assemblies, and (2) within that activated portion, a smaller subset of items in the focus of attention. The activated memory would consist of a fragmented soup of all kinds of activated features (sensory, phonological, orthographic, spatial, and semantic), whereas the focus of attention would contain just a few well-integrated items or chunks.

Contributions of George Miller—Miller (1956) discussed the limitation in how many items can be held in immediate memory. In the relevant test procedure, a list of items is seen or heard and immediately afterward (that is, with no imposed retention interval), the list must be repeated verbatim. The ability to do so was said to be limited to about seven chunks, where a chunk is a meaningful unit. For example, the random digit list *582931* may have to be encoded initially as six chunks, one per digit, whereas the sequence *123654* probably can be encoded by most adults as only two chunks (an ascending triplet followed by a descending triplet). Subsequent work has suggested that the number seven is a practical result that emerges on the basis of strategies that participants use and that, when it is not possible to use chunking or covert verbal rehearsal to help performance, adults typically can retain only 3 or 4 pre-existing chunks (Chen & Cowan, 2009; Cowan, 2001; Cowan, Rouder, Blume, & Saults, 2012; Luck & Vogel, 1997; Rouder et al., 2008).

The first mention I have found of the term working memory comes from a book by Miller, Galanter, and Pribram (1960), *Plans and the structure of behavior*. The title itself, and the concept of organization, seems reminiscent of the earlier work by Hebb (1949), *The organization of behavior*. Miller et al. observed that daily functioning in the world requires a hierarchy of plans. For example, your plan to do well at work requires a sub-plan to be there at time in the morning, which in turn may require sub-plans to eat breakfast, shower, get dressed, gather work materials, and so on. Each of these plans also may have sub-plans, and you may have competing plans (such as choosing an after-work activity, calling your mother, or acquiring food for dinner). Our working memory was said to be the mental faculty whereby we remember the plans and sub-plans. We cannot think about all of them at once but we might, for example, keep in mind that the frying pan is hot while retrieving a knife from the drawer, and we may keep bringing to mind the approximate time so as not to

be late. Working memory was said to be the facility that is used to carry out one sub-plan while keeping in mind the necessary related sub-plans and the master plan.

Contributions of Donald Broadbent—In Great Britain, Broadbent's (1958) book helped to bring the conversation out of the behaviorist era and into an era of cognitive psychology. In a footnote within the book, he sketched a rough information processing diagram that showed information progressing from a sensory type of store that holds a lot of information briefly, through an attention filter to essentially a working memory that holds only a few items, to a long-term memory that is our storehouse of knowledge accumulated through a lifetime. The empirical basis for the model came largely from his work with selective attention, including many dichotic listening studies in which the task was to listen to the message from one ear and ignore the message from the other ear, or report both messages in some order. The motivation for this kind of research came largely from practical issues provoked by World War II, such as how to help a pilot listen to his own air traffic control message while ignoring messages meant for other pilots but presented in the same channel. An important theoretical outcome, however, was the discovery of a difference between a large-capacity but short-lived sensory memory that was formed regardless of attention, and a longer-lived but small-capacity abstract working memory that required attention.

Contributions of Alan Baddeley and Graham Hitch—Miller et al. (1960) may have devised the term working memory but they were not the predominant instigator of the work that has occurred subsequently in the field. Google Scholar does show it with over 5,600 citations. A chapter by Baddeley and Hitch (1974), though, is listed with over 7,400 citations and a 1992 Science article summarizing that approach has over 14,500 citations. In the 1974 chapter, the term working memory was used to indicate a system of temporary memory that is multifaceted, unlike the single store such as James' primary memory, or the corresponding box in Broadbent's (1958) model, or an elaborated version of it as in the model of Atkinson and Shiffrin (1968) model, none of which would do. In fact, a lot of investigators in the 1960's proposed variations of information processing models that included a single short-term memory store, and Baddeley often has referred to these together, humorously, as the "modal model," providing a sketch of it with sensory memory, short-term memory, and long-term memory boxes as in the Broadbent and the Atkinson/ Shiffrin models. (When the humor and the origin of the phrase "modal model" are forgotten, yet the phrase is still widely used, it seems sad somehow.)

The main point emphasized by Baddeley and Hitch (1974) is that there were diverse effects that appeared to implicate short-term memory, but that did not converge to a single component. Phonological processing interfered most with phonological storage, visual-spatial processing interfered with visual-spatial storage, and a working memory load did not seem to interfere much with superior memory for the end of a list, or recency effect. Conceptual learning did not depend heavily on the type of memory that was susceptible to phonological similarity effects, and a patient with a very low memory span was still able to learn new facts. To account for all of the dissociations, they ended up concluding that there was an attention-related control system and various storage systems. These included a

phonological system that also included a covert verbal rehearsal process, and a visual-spatial storage system that might have its own type of non-verbal rehearsal. In the 1974 version of the theory, there were attention limits on the storage of information as well as on processing. In a 1986 book, Baddeley eliminated the attention-dependent storage but in a 2000 paper, a new component was added in the form of an episodic buffer. This buffer might or might not be attention-dependent and is responsible for holding semantic information for the short term, as well as the specific binding or association between phonological and visual-spatial information. Baddeley and Hitch called the assembly or system of storage and processing in service of holding information in an accessible form working memory, the memory one uses in carrying out cognitive tasks of various kinds (i.e., cognitive work).

Model of Cowan (1988)—Through the years, there were several other proposals that alter the flavor of the working memory proposal. Cowan (1988) was concerned with how we represent what we know and do not know about information processing. The "modal models" of which Baddeley has spoken began with Broadbent's (1958) model in which the boxes were shown to be accessed in sequence, comparable to a computer flow chart: first sensory memory, then an attention filter, then short-term memory, and then long-term memory. Atkinson & Shiffrin (1968) preserved the flow chart structure but added recursive entry into the boxes, in the form of the control processes. Baddeley and Hitch (1974) and Baddeley (1986) instead used a processing diagram in which the boxes could be accessed in parallel. One presumably could enter some information into phonological storage while concurrently entering other information into visual-spatial storage, with interacting modules and concurrent executive control.

Cowan (1988, 1995, 1999, 2001, 2005) recoiled a bit from the modules and separate boxes, partly because they might well form an arbitrarily incomplete taxonomy of the systems in the brain. (Where would spatial information about sound go? Where would touch information go? These types of unanswered questions also may have helped motivate the episodic buffer of Baddeley, 2000.) There could be multiple modules, but because we do not know the taxonomy, they were all thrown into the soup of activated long-term memory. Instead of separate boxes, I attempted to model on a higher level at which distinctions that were incomplete were not explicitly drawn into the model, and mechanisms could be embedded in other mechanisms. Thus, there was said to be a long-term memory, a subset of which was in an activated state (cf. Hebb, 1949), and within that, a smaller subset of which was in the focus of attention (cf. James, 1890). Dissociations could still occur on the basis of similarity of features; two items with phonological features will interfere with one another, for example, more than one item with phonological features and another item with only visual-spatial features. The model still included central executive processes.

Compared to Baddeley and Hitch (1974), Cowan (1988) also placed more emphasis on sensory memory. It is true that printed letters, like spoken letters, are encoded with speech-based, phonological features that can be confused with each other in working memory (e.g., Conrad, 1964). Nevertheless, there is abundant other evidence that lists presented in a spoken form are remembered much better, in particular at the end of the list, than verbal lists presented in printed form (e.g., Murdock & Walker, 1969; Penney, 1989).

The attention filter also was internalized in the model of Cowan (1988). Instead of information having to pass through a filter, it was assumed that all information activates long-term memory to some degree. The mind forms a neural model of what it has processed. This will include sensory information for all stimuli but, in the focus of attention, much more semantic information than one finds for unattended information. Incoming information that matches the current neural model becomes habituated, but changes that are perceived cause dishabituation in the form of attentional orienting responses toward the dishabituated stimuli (cf. Sokolov, 1963). Such a system has properties similar to the attenuated filtering model of Treisman (1960) or the pertinence model of Normal (1968). Attention is controlled in this view dually, often with a struggle between voluntary executive control and involuntary orienting responses.

How consistent is Cowan (1988) with the Baddeley and Hitch model? Contributions of Robert Logie—With the addition of the episodic buffer, the model of Baddeley and Hitch makes predictions that are often similar to those of Cowan (1988). There still may be important differences, though. An open question is whether the activated portion of long-term memory of Cowan (1988) functionally serves the same purpose as the phonological and visual-spatial buffers of Baddeley and Hitch (1974) and Baddeley (1986). Robert Logie and colleagues argue that this cannot be, inasmuch as visual imagery and visual short-term memory are dissociated (Borst, Niven, & Logie, 2012; Logie & van der Meulen, 2009; van der Meulen, Logie, & Della Sala, 2009). Irrelevant visual materials interfere with the formation of visual imagery but not with visual storage, whereas tapping in a spatial pattern interferes with visual storage but not the formation of visual images. According to the model that these sources put forward, visual imagery involves activation of long-term memory representations, whereas visual short-term storage is a separate buffer. Although this is a possibility that warrants further research, I am not yet convinced. There could be other reasons for the dissociation. For example, in the study of van der Meulen et al., the visual imagery task involved detecting qualities of the letters presented (curved line or not, enclosed space or not, etc.) and these qualities could overlap more with the picture interference; whereas the visual memory task involved remembering letters in upper and lower case visually, in the correct serial order, and the serial order property may suffer more interference from tapping in a sequential spatial pattern. Testing of the generality of the effects across tasks with different features is needed.

Other models of cross-domain generality—One difference between the Baddeley (1986) framework and that of Cowan (1988) was that Cowan placed more emphasis on the possibility of interference between domains. There has been a continuing controversy about the extent to which verbal and nonverbal codes in working memory interfere with one another (e.g., Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Cowan & Morey, 2007; Fougnie & Marois, 2011; Morey & Bieler, 2013). The domain-general view has extended to other types of research. Daneman and Carpenter (1980) showed that reading and remembering words are tasks that interfere with one another, with the success of remembering in the presence of reading a strong correlate of reading comprehension ability. Engle and colleagues (e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004) showed that this sort of effect does not just occur with verbal materials, but occurs even with

storage and processing in separate domains, such as spatial recall with verbal memory. They attributed individual differences primarily to the processing tasks and the need to hold in mind task instructions and goals while suppressing irrelevant distractions.

Barrouillet and colleagues (e.g., Barrouillet, Portrat, & Camos, 2011; Vergauwe, Barrouillet, & Camos, 2010) emphasized that the process of using attention to refresh items, no matter whether verbal or nonverbal in nature, takes time and counteracts decay. They provided complex tasks involving concurrent storage and processing, like Daneman and Carpenter and like Engle and colleagues. The key measure is cognitive load, the proportion of time that is taken up by the processing task rather than being free for the participant to use to refresh the representations of items to be remembered. The finding of Barrouillet and colleagues has been that the effect of cognitive load on the length of list that can be recalled, or memory span, is a negative linear (i.e., deleterious) effect. They do also admit that there is a verbal rehearsal process that is separate from attentional refreshing, with the option of using either mode of memory maintenance depending on the task demands (Camos, Mora, & Oberauer, 2011), but there is more emphasis on attentional refreshing than in the case of Baddeley and colleagues, and the approach therefore seems more in keeping with Cowan (1988) with its focus of attention (regarding refreshing see also Cowan, 1992).

Ongoing controversies about the nature of working-memory memory limits—
There are theoretically two basic ways in which working memory could be more limited than long-term memory. First, It could be limited in terms of how many items can be held at once, a capacity limit that Cowan (1998, 2001) tentatively ascribes to the focus of attention. Second, it could be limited in the amount of time for which an item remains in working memory when it is no longer rehearsed or refreshed, a decay limit that Cowan (1988) ascribed to the activated portion of long-term memory, the practical limit being up to about 30 seconds depending on the task.

Both of these limits are currently controversial. Regarding the capacity limit, there is not much argument that, within a particular type of stimulus coding (phonological, visual-spatial, etc.), normal adults are limited to about 3 or 4 meaningful units or chunks. The debate is whether the limit occurs in the focus of attention, or because materials of similar sorts interfere with one another (e.g., Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). In my recent, still-unpublished work, I suggest that the focus of attention is limited to several chunks of information, but that these chunks can be off-loaded to long-term memory and held there, with the help of some attentional refreshing, while the focus of attention is primarily used to encode additional information.

Regarding the memory loss or decay limit, some studies have shown no loss of information for lists of printed verbal materials across periods in which rehearsal and refreshing have apparently been prevented (Lewandowsky, Duncan, & Brown, 2004; Oberauer & Lewandowsky, 2008). Nevertheless, for arrays of unfamiliar characters followed by a mask to eliminate sensory memory, Ricker and Cowan (2010) did find memory loss or decay (cf. Zhang & Luck, 2009). In further work, Ricker et al. (in press) suggested that the amount of decay depends on how well the information is consolidated in working memory (cf. Jolicoeur & Dell'Acqua, 1998). Given that the time available for refreshing appeared to be

inversely related to the cognitive load, the consolidation process that seems critical is not interrupted by a mask but continues after it. This consolidation process could be some sort of strengthening of the episodic memory trace based on attentional refreshing in the spirit of Barrouillet et al. (2011). If so, the most important effect of this refreshing would not be to reverse the effects of decay temporarily, as Barrouillet et al. proposed, but rather to alter the rate of decay itself. Our plans for future research include investigation of these possibilities.

Long-term working memory—It is clear that people function quite well in complex environments in which detailed knowledge must be used in an expert manner, despite a severe limit in working memory to a few ideas or items at once. What is critical in understanding this paradox of human performance is that each slot in working memory can be filled with a concept of great complexity, provided that the individual has the necessary knowledge in long-term memory. This point was made by Miller (1956) in his concept of combining items to form larger chunks of information, with the limit in working memory found in the number of chunks, not the number of separate items presented for memorization. Ericsson and Kintsch (1995) took this concept further by expanding the definition of working memory to include relevant information in long-term memory.

Although we might quibble about the best definition of working memory, it seems undeniable that long-term memory is often used as Ericsson and Kintsch (1995) suggest. An example is what happens when one is holding a conversation with a visitor that is interrupted by a telephone call. During the call, the personal conversation with one's visitor is typically out of conscious working memory. After the call, however, with the visitor serving as a vivid cue, it is often possible to retrieve a memory of the conversation as a recent episode and to remember where this conversation left off. That might not be possible some days later. This use of long-term memory to serve a function similar to the traditional working memory, thus expanding the person's capabilities, was termed long-term working memory by Ericsson and Kintsch. Cowan (1995) alluded to a similar use of long-term memory for this purpose but, not wanting to expand the definition of working memory, called the function virtual short-term memory, meaning a use of long-term memory in a way that short-term memory is usually used. It is much like the use of computer memory that allows the computer to be turned off in hibernation mode and later returned to its former state when the memory is retrieved.

Given the ability of humans to use long-term memory so adeptly, one could ask why we care about the severe working memory capacity limit at all. The answer is that it is critical when there is limited long-term knowledge of the topic. In such circumstances, the capacity of working memory can determine how many items can be held in mind at once in order to use the items together, or to link them to form a new concept in long-term memory. This is the case in many situations that are important for learning and comprehension. One simple example of using items together is following a set of instructions, e.g., to a preschool child, *put your drawing in your cubby and then go sit in the circle*. Part of that instruction may be forgotten before it is carried out and teachers must be sensitive to that possibility. A simple example of linking items together is in reading a novel, when one listens to a description of a character and melds the parts of the description to arrive at an overall personality sketch that can be formed in long-term memory. Inadequate use of working memory during reading

may lead to the sketch being incomplete, as some descriptive traits are inadvertently ignored. Knowledge of this working memory limit can be used to improve one's writing by making it easier to remember and comprehend.

Paas and Sweller (2012) bring up the distinction between biologically primary and secondary knowledge (Geary, 2008) and suggest (p. 29) that "Humans are easily able to acquire huge amounts of biologically primary knowledge outside of educational contexts and without a discernible working memory load." Examples they offered were the learning of faces and learning to speak. It may well be the case that individual faces or spoken words quickly become integrated chunks in long-term memory (and, I would add, the same seems true for objects in domains of learned expertise, e.g., written words in adults). Nevertheless, the biologically-primary components are used in many situations in which severe capacity limits do apply. In these situations, the added memory demand is considered biologically secondary. An example is learning which face should be associated with which name. If four novel faces are shown on a screen and their names are vocally presented, these name-face pairs cannot be held in working memory at once, so it is difficult to retain the information and it often takes additional study of one pair at a time to remember the name-face pairing.

Specific mathematical models—Here I have been selective in examining models of working memory that are rather overarching and verbally specified. By limiting the domain of applicability and adding some processing assumptions, other researchers throughout the years have been able to formulate models that make mathematical predictions of performance in specific situations. We have learned a lot from them but they are essentially outside of the scope of this review given limited space and given my own limitations. For examples of such models see Brown, Neath, & Chater, 2007; Burgess & Hitch, 1999; Cowan et al., 2012; Farrell & Lewandowsky, 2002; Hensen, 1998; Murdock, 1982; Oberauer & Lewandowsky, 2011). The importance of these models is that they make clear the consequences of our theoretical assumptions. In order to make quantitative predictions, each mathematical assumption must be made explicit. It is sometimes found that the effects of certain proposed mechanisms, taken together, are not what one might assume from a purely verbal theory. Of course, some of the assumptions that one must make to eke out quantitative predictions may be unsupported, so I believe that the best way forward in the field is to use general verbal, propositional thinking some of the time and specific quantitative modeling other times, working toward a convergence of these methods toward a common theory.

Summary: Status of Working Memory

The progress in this field might be likened to an upward spiral. We make steady progress but meanwhile, we go in circles. The issues of the nature of working memory limits have not changed much from the early days. Why is the number of items limited? Why is the duration limited? What makes us forget? How is it related to the conscious mind and to neural processes? These questions are still not answered. At the same time, we have agreement about what can be found in particular circumstances. Set up the stimuli one way, and there is interference between modalities. Set it up another way and there appears to be much less interference. Set it up one way and items are lost rapidly across time. Set it up a different

way, and there is much less loss. There are brain areas associated with the focus of attention and with working memory across modalities (Cowan, 2011; Cowan, Li et al., 2011; Todd & Marois, 2004; Xu & Chun, 2006). This is progress awaiting an adequate unifying theory.

What we do know has practical implications. To avoid overtaxing an individual's working-memory capabilities, one should avoid presenting more than a few items or ideas at once, unless the items can be rapidly integrated. One should also avoid making people hold on to unintegrated information for a very long time. For example, I could write a taxing sentence like, *It is said that, if your work is not overwhelming, your car is in good repair, and the leaves have changed color, it is a good time for a fall vacation*. However, that sentence requires a lot from the reader's working memory. I could reduce the working memory load by not making you wait for the information that provides the unifying theme, keeping the working memory load low: *It is said that a good time for a fall vacation is when your work is not overwhelming, your car is in good repair, and the leaves have changed color*.

Working Memory and Cognitive Development

There is no question that working-memory capabilities increase across the life span of the individual. In early tests of maturation (e.g., Bolton, 1892), and to this day in tests of intelligence, children have been asked to repeat lists of random digits. The length of list that can be successfully repeated on some predefined proportion of trials is the digit span. It increases steadily with childhood maturation, until late childhood. When the complexity of the task is increased, the time to adult-like performance is extended a bit further, with steady improvement throughout childhood (for an example see Gathercole, Pickering, Ambridge, & Wearing, 2004).

As we saw in the introductory section, clear practical findings do not typically come with a clear understanding of the theoretical explanation. There have been many explanations over the years for the finding of increasing memory span with age (e.g., see Bauer & Fivush, in press; Courage & Cowan, 2009; Kail, 1990). These explanations may lead to differing opinions of the best course for learning and education, as well.

Explanations Based on Capacity

Explanations of intellectual growth based on working memory capacity stem from what has been called the neoPiagetian school of thought. Jean Piaget outlined a series of developmental stages, but with no known underlying reason for the progression between stages. Pascual-Leone and Smith (1969) attributed the developmental increases to increases in the number of items that could be held in mind at once.

The theory becomes more explicit with the contributions of Halford, Phillips, and Wilson (1998) and Andrews and Halford (2002). They suggests that it is the number of associations between elements that is restricted and that this matters because it limits the complexity of thought. In my example above, the concept of a tiger versus lion versus zebra requires concurrent consideration of the animal's shape and presence or absence of stripes. Similarly, addition requires the association between three elements: the two elements being added and the sum. A concept like *bigger than* is a logical relation requiring three slots, e.g., *bigger*

than (dog, elephant). Ratios require the coordination of four elements (e.g., 4/6 is equivalent to 6/9) and therefore are considerably harder to grasp, according to the theory (see Halford, Cowan, & Andrews, 2007).

This concept is quite promising and might even appear to be "the only game in town" when it comes to trying to understand the age limits on children's ability to comprehend ideas of various levels of complexity. One problem with it is that it is not always straightforward to determine the *arity* of a concept, or number of ideas that must be associated. For example, a young child might understand the concept *big(elephant)* and then might be able to infer that elephants are bigger than dogs, without being able to use the concept of *bigger than* in a consistent manner more generally. The concept from Miller (1956) that items can be combined using knowledge to form larger chunks also applies to associations, and it is not clear how to be sure that the level of complexity actually is what it is supposed to be. Knowledge allows some problems to be solved with less working memory requirement.

Explanations Based on Knowledge

It is beyond question that knowledge increases with age. Perhaps this knowledge increase is the sole reason for developmental change in working memory, it has been argued. Chi (1978) showed that children with an expertise in the game of chess could remember chess configurations better than adults with no such expertise. The expert children presumably could form larger chunks of chess pieces, greatly reducing the memory load. Case, Kurland, and Goldberg (1982) gave adults materials that were unfamiliar and found that both the speed of item identification and the memory span for those materials closely resembled what was found for 6-year-olds on familiar materials. The implication was that the familiarity with the materials determines the processing speed, which in turn determines the span.

Explanations Based on Processing Speed and Strategies

Case et al. (1982) talked of a familiarity difference leading to a speed difference. Others have suggested that, more generally, speed of processing increases with age in childhood and decrease again with old age (e.g., Kail & Salthouse, 1994). This has led to accounts of working memory improvement based on an increased rate of covert verbal rehearsal (Hulme & Tordoff, 1989) or increased rate of attentional refreshing (Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009; Camos & Barrouillet, 2011). At the lower end of childhood, it has been suggested on the basis of various evidence that young children do not rehearse at all (Flavell, Beach, & Chinsky, 1966; Garrity, 1975; Henry, 1991) or do not rehearse in a sufficiently sophisticated manner that is needed to assist in recall (Ornstein & Naus, 1978). When rehearsal aloud is required, the result suggest that the most recently rehearsed items are recalled best (Tan & Ward, 2000).

This view that rehearsal is actually important has been opposed recently. It is not clear that rehearsal must be invoked to explain performance (Jarrold & Citroën, 2013) and if rehearsal takes place, it is not clear exactly what the internal processes are (e.g., cumulative repetition of the list? Repetition of each item as it is presented?).

In the case of using attention to refresh information, an interesting case can be made. Children who are too young (about 4 years of age and younger) do not seem to use attention

to refresh items. For them, the limit in performance depends on the duration of the retention interval. For older children and adults, who are able to refresh, it is not the absolute duration but the cognitive load that determines performance (Barrouillet et al., 2011). The "phase change" in performance that is observed here with the advent of refreshing is perhaps comparable to the phase change that is seen with the advent of verbal rehearsal (Henry, 1991), though the evidence may be stronger in the case of refreshing.

Re-assessment of Capacity Accounts

We have seen that there are multiple ways in which children's working memory performance gets better with maturity. There are reasons to care about whether the growth of capacity is primary, or whether it is derived from some other type of development. For example, if the growth of capacity results only from the growth of knowledge, then it should be possible to teach any concept at any age, if the concept can be made familiar enough. If capacity differences come from speed differences, it might be possible to allow more time by making sure that the parts to be incorporated into a new concept are presented sufficiently slowly.

We have done a number of experiments suggesting that there is something to capacity that changes independent of these other factors. Regarding knowledge, relevant evidence was provided by Cowan, Nugent, Elliott, Ponomarev, and Saults (1999) in their test of memory for digits that were unattended while a silent picture-rhyming game was carried out. The digits were attended only occasionally, when a recall cue was presented about 1 s after the last digit. The performance increase with age throughout the elementary school years was just as big for small digits (1, 2, 3), which are likely to be familiar, as for large digits (7, 8, 9), which are less familiar. Gilchrist, Cowan, and Naveh-Benjamin (2009) further examined memory for lists of unrelated, spoken sentences in order to distinguish between a measure of capacity and a measure of linguistic knowledge. The measure of capacity was an access rate, the number of sentences that were at least partly recalled. The measure of linguistic knowledge was a completion rate, the proportion of a sentence that was recalled, provided that at least part of it was recalled. This sentence completion rate was about 80% for both first and sixth grader children, suggesting that for these simple sentences, there was no age difference in knowledge. Nevertheless, the number of sentences accessed was considerably smaller in first-grade children than in sixth-grade children (about 2.5 sentences vs. 4 sentences). I conclude, tentatively at least, that knowledge differences cannot account for the age difference in working memory capacity.

We have used a different procedure to help rule out a number of factors that potentially could underlie the age difference in observed capacity. It is based on a procedure that has been well-researched in adults (Luck & Vogel, 1997). On each trial of this procedure, an array of simple items (such as colored squares) is presented briefly and followed by a retention interval of about 1 s, and then a single probe item is presented. The latter is to be judged identical to the array item from the same location, or a new item. This task is convenient partly because there are mathematical ways to estimate the number of items in working memory (Cowan, 2001). If k items are in working memory and there are N items in the array, the likelihood that the probed item is known is k/N, and a correct response can

also come from guessing. It is possible to calculate k, which for this procedure is equal to $N(h\text{-}f)$, where h refers the proportion of change trials in which the change was detected (hits) and f refers to the proportion of no-change trials in which a change was incorrectly reported (false alarms).

One possibility is that younger children remember less of the requested information because they attend to more irrelevant information, cluttering working memory (for adults, cf. Vogel, McCollough, & Machizawa, 2005). To examine this, Cowan, Morey, AuBuchon, Zwilling, and Gilchrist (2010) presented both colored circles and colored triangles and instructed participants to pay closer attention to one shape, which was tested on 80% of the trials in critical blocks. When there were 2 triangles and 2 circles, memory for the more heavily-attended shape was better than memory for the less-attended shape, to the same extent in children in Grades 1–2 and Grades 6–7, and in college students. Yet, the number of items in working memory was much lower in children in Grades 1–2 than in the two older groups. It did not seem that the inability to filter out irrelevant information accounted for the age difference in capacity.

Another possibility is that in Cowan et al. (2010), the array items occurred too fast for the younger children to encode correctly. To examine this, Cowan, AuBuchon, Gilchrist, Ricker, & Saults (2011) presented the items one at a time at relatively slow, a 1-item-per-second rate. The results remained the same as before. In some conditions, the participant had to repeat each color as it was presented or else say "wait" to suppress rehearsal; this articulatory manipulation, too, left the developmental effect unchanged. It appears that neither encoding speed nor articulation could account for the age differences. So we believe that age differences in capacity may be primary rather than derived from another process.

Age differences in capacity still could occur because of age differences in the speed of a rapid process of refreshment, and from the absence of refreshment in young children (Camos & Barrouillet, 2011). Alternatively, it could occur because of age differences in some other type of speed, neural space, or efficiency. This remains to be seen but at least we believe that there is a true maturational change in working memory capacity underlying age differences in the ability to comprehend materials of different complexity. This is in addition to profound effects of knowledge acquisition and the ability to use strategies.

The use of strategies themselves may be secondary to the available working memory resources to carry out those strategies. According to the neoPiagetian view of Pascual-Leone and Smith (1969), for example, the tasks themselves share resources with the data being stored. Cowan et al. (2010) found that when the size of the array to be remembered was large (3 more-relevant and 3 less-relevant items, rather than 2 of each) then young children were no longer able to allocate more attention to the more-relevant items. The attentional resource allocated to the items in the array was apparently deducted from the resource available to allocate attention optimally.

In practical terms, it is worth remembering that several aspects of working memory are likely to develop: capacity, speed, knowledge, and the use of strategies. Although it is not

always easy to know which process is primary, these aspects of development all should contribute in some way to our policies regarding learning and education.

Working Memory and Learning

In early theories of information processing, up through the current period, working memory was viewed as a portal to long-term memory. In order for information to enter long-term memory in a form that allows later retrieval, it first must be present in working memory in a suitable form. Sometimes that form appears modality-specific. For example, Baddeley, Papagno, and Vallar (1988) wondered how it could be that a patient with a very small verbal short-term memory span, 2 or 3 digits at most, could function so well in most ways and exhibit normal learning capabilities. The answer turned out to be that she displayed a very selective deficit: she was absolutely unable to learn new vocabulary. This finding led to a series of developmental studies showing that individual differences in phonological memory are quite important for differences in word-learning capability in both children and adults (Baddeley, Gathercole, & Papagno, 1998; Gathercole & Baddeley, 1989, 1990).

Aside from this specific domain, there are several ways in which working memory can influence learning. It is important to have sufficient working memory for concept formation. The control processes and mnemonic strategies used with working memory also are critical to learning.

Working Memory and Concept Formation

Learning might be thought of in an educational context as the formation of new concepts. These new concepts occur when existing concepts are joined or bound together. Some of this binding is mundane. If an individual knows what the year 1776 means and also what the Declaration of Independence is (at least in enough detail to remember the title of the declaration), then it is possible to learn the new concept that the Declaration of Independence was written in the year 1776. Other times, the binding of concepts may be more interesting and there may be a new conceptual leap involved. For example, a striped cat is a tiger. As another simple example, to understand what a parallelogram is, the child has to understand what the word *parallel* means, and further to grasp that two sets of parallel lines intersect with one another. The ideas presumably must co-exist in working memory for the concept to be formed.

For the various types of concept formation, then, the cauldron is assumed to be working memory. According to my own view, the binding of ideas occurs more specifically in the focus of attention. We have taken a first step toward verifying that hypothesis. Cowan, Donnell, and Saults (in press) presented lists of words with an incidental task: to report the most interesting word in each presented list. Later, participants completed a surprise test in which they were asked whether pairs of words came from the same list; the words were always one or two serial positions apart in their respective lists, but sometimes were from the same list and sometimes from different lists. The notion was that the link between the words in the same list would be formed only if the words had been in the focus of attention at the same time, which was much more likely for short lists than for long lists. In keeping with this hypothesis, performance was better for words from short lists of 3 items (about

59%) than for words from lists of 6 or 9 items (about 53%). This is a small effect, but it is still important that there was unintentional learning of the association between items that were together in the focus of attention just once, when there was no intention of learning the association.

The theory of Halford et al. (1998) may be the best articulated theory suggesting why a good working memory is important for learning. (In this discussion, a "good" working memory is simply one that can keep in mind sufficient items and their relations to one another to solve the problem at hand, which may require a sufficient combination of capacity, speed, knowledge, and available strategies.) More complex concepts require that one consider the relationship between more parts. A person's working memory can be insufficient for a complex concept. It may be possible to memorize that concept with less working memory, but not truly to understand the concept and work with it. Take, for example, use of the concept of transitivity in algebra. If $a+b=c+d$ and $c+d=e$, then we can conclude that $a+b=e$ because equality is transitive. Yet, a person who understands the rules of algebra still would not be able to draw the correct inference if that person could not concurrently remember the two equations. Even if the equations are side by side on the page, that does not mean that they necessarily can be encoded into working memory at the same time, which is necessary in order to draw the inference. Lining up the equations vertically for the learner and then inviting the learner to apply the rule by rote is a method that can be used to reduce the working memory load, perhaps allowing the problem to be solved. However, working out the problem that way will not necessarily produce the insight needed to set up a new problem and solve it, because setting up the problem correctly requires the use of working memory to understand what should be lined up with what. So if the individual does not have sufficient working memory capacity, a rote method of solution may be helpful for the time being. More importantly, though, the problem could be set up in a more challenging manner so that the learner is in the position of having to use his or her working memory to store the information. By doing so, the hope is that successful solution of the problem then will result in more insight that allows the application of the principles to other problems. That, in fact, is an expression of the issues that may lead to the use of word problems in mathematics education.

Working Memory and Control Processes

Researchers appear to be in fairly good agreement that one of the most important aspects of learning is staying on task. If one does not stick to the relevant goals, one will learn something perhaps, but it will not be the desired learning. Individuals who test well on working memory tasks involving a combination of storage and processing have been shown to do a better job staying on task.

A good experimental example of how staying on task is tied to working memory is one carried out by Kane and Engle (2003) using a well-known task designed long ago by John Ridley Stroop. In the key condition within this task, one is to name the color of ink in which color words are written. Sometimes, the color of ink does not match the written color and there is a tendency to want to read the word instead of naming the color. This effect can be made more treacherous by presenting stimuli in which the word and color match on most

trials, so that the participant may well lapse into reading and lose track of the correct task goal (naming the color of ink). What that happens, the result is an error or long delay on the occasional trials for which the word and ink do not match. Under those circumstances, the individuals who are more affected by the Stroop conditions are those with relatively low performance on the operation span test of working memory (carrying out arithmetic problems while remembering words interleaved with those problems).

In more recent work, Kane et al. (2007) has shown that low-span individuals have more problems attending in daily life. Participants carried devices that allowed them to respond at unpredictable times during the day, reporting what they were doing, what they wanted to be doing, and so on. It was found that low-span individuals were more likely to report that their minds were wandering away from the tasks on which they were trying to focus attention. This, however, did not occur on all tasks. The span-related difference in attending was only for tasks in which they reported that they wanted to pay attention. When participants reported that they were bored and did not want to pay attention, mind-wandering was just as prevalent for high spans as for low spans.

Although this work was done on adults, it has implications for children as well. Gathercole, Lamont, and Alloway (2006) suggest that working memory failures appear to be a large part of learning disabilities. Children who were often accused of not trying to follow directions tested out as children with low working memory ability. They were often either not able to remember instructions or not able to muster the resources to stick to the task goal and pay attention continually, for the duration needed. Children with various kinds of learning and language disability generally test below grade level on working memory procedures, and children with low working memory and executive function don't do well in school (e.g., Sabol & Pianta, 2012).

Of course, central executive processes must do more than just maintain the task goal. The way in which information is converted from one form to another, the vigilance with which the individual searches for meaningful connections between elements and new solutions, and self-knowledge about what areas are strong or weak all probably play important roles in learning.

Working Memory and Mnemonic Strategies

There also are special strategies that are needed for learning. For example, a sophisticated rehearsal strategy for free recall of a list involves a rehearsal method that is cumulative. If the first word on the list is a cow, the second is a fish, and the third a stone, one ideally should rehearse cumulatively: *cow…cow, fish….cow, fish, stone…* and so on (Ornstein & Naus, 1978). Cowan, Saults, Winterowd, and Sherk (1991) showed that young children did not carry out cumulative rehearsal the way older children do and could not easily be trained to do so, but that their memory improved when cumulative rehearsal was overtly supported by cumulative presentation of stimuli.

For long-term learning, maintenance rehearsal is not nearly as effective a strategy as elaborative rehearsal, in which a coherent story is made on the basis of the items; this takes

time but results in richer associations between items, enhancing long-term memory provided that there is time for it to be accomplished (e.g., Craik & Watkins, 1973).

In addition to verbal and elaborative rehearsal, Barrouillet and colleagues (2011) have discussed attentional refreshing as a working-memory maintenance process. We do not yet know what refreshing looks like on a moment-to-moment basis or what implications this kind of maintenance strategy has for long-term learning. It is a rich area for future research.

The most general mnemonic strategy is probably chunking (Miller, 1956), the formation of new associations or recognition of existing ones in order to reduce the number of independent items to keep track of in working memory. The power of chunking is seen in special cases in which individuals have learned to go way beyond the normal performance. Ericsson, Chase, and Faloon (1980) studied an individual who learned, over the course of a year, to repeat lists of about 80 digits from memory. He learned to do so starting with a myriad of athletic records that he knew so that, for example, 396 might be recoded as a single unit, 3.96 minutes, a fairly fast time for running the mile. After applying this intensive chunking strategy in practice for a year, a list of 80 digits could be reduced to several sub-lists, each with associated sub-parts. The idea would be that the basic capacity has not changed but each working-memory slot is filled with quite a complex chunk. In support of this explanation, individuals of this sort still remain at base level (about 7 items) for lists of items that were not practiced in this way, e.g., letters. (For a conceptual replication see Ericsson, Delaney, Weaver, & Mahadevan, 2004; Wilding, 2001)

Although we cannot all reach such great heights of expert performance, we can do amazing things using expertise. For example, memorization of a song or poem is not like memorization of a random list of digits because there are logical connections between the words and between the lines. A little working memory then can go a long way.

The importance of a good working memory comes in when something new is learned, and logical connections are not yet formed so the working memory load is high. When there are not yet sufficient associations between the elements of a body of material, working memory is taxed until the material can be logically organized into a coherent structure. Working memory is thought to correlate most closely with fluid intelligence, the type of intelligence that involves figuring out solutions to new problems (e.g., Wilhelm & Engle, 2005). However, crystallized intelligence, the type of intelligence that involves what you know, also is closely related to fluid intelligence. The path I suggest here is that a good working memory assists in problem-solving (hence fluid intelligence); fluid intelligence and working memory then assist in new learning (hence crystallized intelligence).

Working Memory and Education

We have sketched the potential relation between working memory and learning. How is that to be translated into lessons for education? There is a large and diverse literature on this topic. As a starting point to illustrate this diversity, I will describe the chapters chosen for the book, *Working memory and education* (Pickering, 2006). After an introductory chapter on working memory (A. Baddeley), the book includes two chapters on the relation between working memory and reading (one by P. de Jong and another by K. Cain). There is a chapter

on the relation between working memory and mathematics education (R. Bull and K.A. Espy), learning disabilities (H.L. Swanson), attention disorders (K. Cornish and colleagues), and deafness (M. Keehner & J. Atkinson). Other chapters cover more general topics, including the role of working memory in the classroom (S. Gathercole and colleagues), the way to assess working memory in children (S. Pickering), and sources of working memory deficit (M. Minear and P. Shah). It is clear that many avenues of research relate working memory to education, and I cannot travel along all of them in this review.

To organize a diverse field, what I can do is to distinguish between several different basic approaches have been tried. First, one can try to teach to the level of the learner's working memory. The points described in the article up to this point should be kept in mind when one is trying to discern and understand what a particular learner can and cannot do. Second, one can try to use training exercises to improve working memory, which, investigators have hoped, would allow a person to be able to learn more and solve problems more successfully. The message I would give here is to be wary, given the rudimentary state of the evidence in a difficult field and the plethora of companies selling working memory training exercises. Third, one might contemplate the role of working memory for the most critical goals of education, in a broad sense. These topics will be examined one at a time.

Teaching to the Level of Working Memory

The classic adaptation of education to cognitive development and the needs of learning has been to try to adjust the materials to fit the learner. For example, there has been considerable discussion of the need to delay teaching concepts of arithmetic at least until the children understand the basic underlying concept of one-to-one correspondence; that is, the idea that there are different numbers in a series and that each number is assigned to just one object, in order to count the objects (e.g., Gelman, 1982). Halford et al. (2007) provide rough description of what complexity of concepts to expect for each age range, based on working-memory limits (see also Pascual-Leone & Johnson, 2011).

There also are individual differences within an age group in ability that affect how the materials are processed. For example, individuals lower in working memory may prefer to take in information using a verbatim, shallow, or surface processing strategy, rather than try to extract the gist (for one relevant investigation, albeit with mixed results, see Kyndt, Cascallar, & Dochy, 2012). The enjoyment of technological presentations may be greater in students with better abilities in the most relevant types of working memory (e.g., Garcia, Nussbaum, & Preiss, 2011). I would note that the educational enterprise requires that the teacher must decide whether it is best to allow the learner to use a favored strategy, which may be influenced by the student's ability level, or whether it is possible in some cases to instill a more effective strategy even if it does not come naturally to the student.

Sweller and colleagues (Sweller, 2011; Sweller, van Merrienboer, & Paas, 1998) have summarized a body of research literature and a theory about the role of cognitive load in learning and education. Their cognitive load theory is "a theory that emphasizes working memory constraints as determinants of instructional design effectiveness" (Sweller et al., 1998). The theory distinguishes between an intrinsic cognitive load that comes from material to be learned and an extraneous cognitive load that should be kept small enough that the

cognitive resources of the learner are not overly depleted by it. The theory is importantly placed in an evolutionary framework that I will not describe (though above I mentioned the theory's incorporation of the distinction between biologically primary and secondary information). This theory has the advantage of being rather nuanced in that many ramifications of cognitive load are considered. With too high a cognitive load, one runs the risk of the student not being able to follow the presentation, whereas with too low a cognitive load, one runs the risk of insufficient engagement. In future, it might be possible to refine the predictions for classroom learning by combining cognitive load theory with theories of cognitive development, which make some specific predictions about how much capacity is present at a particular age in childhood (e.g., Halford et al., 2007). For further discussion of the theory as applied specifically to multimedia, see Schüler, Scheiter, and Genuchten (2011). Issues arise as to how printed items are encoded (visually, verbally, or both) and how much the combination of verbal and visual codes in multimedia should be expected to tax a common, central cognitive resource and therefore interfere with one another, even when they are intended to be synergic. Both in cognitive psychology and in education, these are key issues currently under ongoing investigation.

An advantage of multimedia and computerized instruction is the possibility of adjusting the instruction to the student's level. This might be done partly on the basis of success; if the student succeeds, the materials can be made more challenging whereas, if the student fails, the materials can be made easier. One potential pitfall to watch for is that, while some students will want to press slightly beyond their zone of comfort and will learn well, others will want an easy time, and may choose to learn less than they would be capable of learning. One way to cope with these issues is through computerized instruction, but with a heavy dose of personal monitoring and adjustment to make sure that the task is sufficiently motivating for every student.

One factor that makes it difficult to teach to the students effectively is that the working memory demands of language production do not always match the demands of the recipients' language comprehension. Consequently, when one is speaking or writing for didactic purposes, one must be careful to consider not only one's own working memory needs, but also those of the listener or reader. There are several obstacles in this regard. Slevc (2011) showed that speakers tend to blurt out what is most readily available in working memory. He used situations that were to be described verbally by the participant, e.g., *A pirate gave a book to the monk*. If one piece of information had already been presented, it was more likely to be described first. For example, if the monk had been presented already but not the book, the participant was more likely to phrase the description differently, as *A pirate gave the monk a book*. This assignment of priority to given information is generally appropriate, given that the speaker and listener (or writer and reader) share the same given information. In this case, though, Slevc shows that the tendency to describe given information first was diminished when the speaking participant was under a working memory load. In a didactic situation such as giving a lecture, it thus seems plausible that the memory load inherent in the situation (remembering and planning what one wants to say in the coming segments of a lecture) may cause the lecturer sometimes to use awkward grammatical structure. Moreover, as mentioned above, learning to speak or write well requires that one bear in mind possible difference between what one

knows as the speaker (or writer) and what the listener (or reader) knows at key moments. For example, if one says, "Marconi was the inventor of the modern radio," then, by the time the full topic of the sentence is known, the name is most likely no longer in the listener's or student's working memory. If, however, one says, "The modern radio was invented by a man named Marconi," the context is set up first, making it easier to retain the name. Bearing in mind what the listener or reader knows and does not yet know is likely to be important both for educators in their own speaking and writing, and also in order to teach students how to speak and write effectively.

Working Memory Training

A much more controversial approach is to use training regimens to improve working memory, thereby improving performance on the educational learning tasks that require working memory (e.g., Klingberg, 2010). It is controversial partly because many people have spent a great deal of money purchasing such training programs before the scientific community has reached an agreement about the efficacy of such programs.

Doing working memory training studies is not easy. One needs a control group that is just as motivated by the task as the training group but without the working memory training aspect. The training task must be adaptive (with rewards for performance that continues to improve with training) and a non-adaptive control group does not adequately control arousal and motivation. Some task that is adaptive but involves long-term learning instead of working memory training may be adequate. Several large-scale reviews and studies have suggested that working memory training sometimes improves performance on the working memory task that is trained, but does not generalize to reasoning tasks that must rely on working memory (in adults, Redick et al., 2013, and Shipstead, Redick, & Engle, 2012; in children, Melby-Lervåg & Hulme, 2013). In somewhat of a contrast, other reviews suggest that the training of executive functions (inhibiting irrelevant information, updating working memory, controlling attention, etc.) does extend at least to tasks that use similar processes (Diamond & Lee, 2011) and some basically concur also for working memory (Chein & Morrison, 2010). So there is an ongoing controversy, even among those who have written meta-analyses and reviews of research.

One might ask how it is possible to improve working memory without having the effect of improving performance on other tasks that rely on working memory. This can happen because there are potentially two ways in which training can improve task performance. First, working memory training theoretically might increase the function of a basic process, much as a muscle can be strengthened through practice. (Or at least, individuals might learn that through diligent exertion of their attention and effort, they can do better.) That is presumably the route hoped for in training of working memory or executive function. Second, though, it is possible for working memory training to result in the discovery of a strategy for completing the task that is better than the strategy used initially. This can improve performance on the task being trained, but the experience and the strategy learned may well be irrelevant to performance on other educational tasks, even those that rely on working memory. This route might be expected if, as I suspect, participants typically look for a way to solve a problem that is not very attention-demanding, unless the payoff is high.

If there is successful working memory training, another issue is whether training is capable of producing super-normal performance or whether it is mostly capable of rectifying deficiencies. By way of analogy, consider physical exercise. If a person is already walking 6 miles a day, there might be little benefit to the heart of adding aerobic exercise. Similarly, if a person is already highly engaged in the environment and using attention control often and effectively during the day, there might be little benefit to the brain of adding working memory exercises. It remains quite conceivable, though, that such exercises are beneficial to certain individuals who are under-utilizing working memory. Nevertheless, as Diamond and Lee (2011) points out, there might be social or emotional reasons why this is the case and such factors would need to be addressed along with, or in some cases instead of, working memory training per se.

Working Memory and the Ultimate Goals of Education

What is the difference between learning and education? This is a question that has long been asked (for a history of the early period of educational psychology in the United States, for example, see Hall, 2003). Do children learn better when they are fed the information intensively, or allowed to explore the material? Should all children be expected to learn the same material, or should children be separated into different tracks and taught the information that is thought to help them the most in their own most likely future walks of life?

A fundamental difference between learning and education, many would agree, is that education should facilitate the acquisition of skills that will promote continued learning after the student leaves school. Of course, after the student leaves school, a major difference is that there is no teacher to decide what is to be learned, or how. Therefore, what seems to be most important, many would agree, is critical thinking skills. There is some sentiment that these skills can be trained (although for an opposing view see Tricot & Sweller, in press). For example, Halpern (1998 p. 449) suggests the following emphases for training critical thinking: "(a) a dispositional component to prepare learners for efforiful cognitive work, (b) instruction in the skills of critical thinking, (c) training in the structural aspects of problems and arguments to promote transcontextual transfer of critical-thinking skills, and (d) a metacognitive component that includes checking for accuracy and monitoring progress toward the goal." Although I could find few well-controlled, peer-reviewed studies supporting the notion that it is possible to train critical thinking skills, optimistic evidence is beginning to roll in. For example, Shim and Walczak (2012) found that professors asking challenging questions resulted in more improvement in both subjective and objective measures of critical thinking. The objective measure required that students clarify, analyze, evaluate, and extend arguments, and increased 0.55 standardized units for every 1-unit increase in the rating of challenging questions asked. The gain was much stronger in students with high pretest scores in critical thinking. Halpern et al. (2012) have designed a computerized module to train critical thinking skills and obtained very encouraging initial results, with well-controlled training experiments in progress according to the report.

One can then ask, to what extent is the training of these higher-level skills dependent on the student's working memory ability? The association is likely to be substantial, given the high

correlation between working memory and reasoning ability even among normal adults (Kyllonen & Christal, 1990; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). There is the possibility that training working memory will in some way improve reasoning and vice versa, though most would agree at this point that the case has not yet been completely made (e.g., Jaeggi & Buschkuehl, 2013; Shipstead et al., 2012).

A current interest of mine is to understand how fallacies in reasoning might be related to fallacies in working memory performance. There appear to be some similarities between the two. One of the best-known reasoning fallacies is confirmation bias. In a key example (Wason & Shapiro, 1971) participants are given a set of cards laid on the table, each having a letter on one side and a number on the other, and are asked which cards must be turned over to assess a rule (e.g., *If a card has a vowel on one side, it has an even number on the other side*). Participants get that they must turn over the cards that can either confirm or disconfirm the rule (in the example, the cards showing vowels). They often fail to realize that they must also turn over cards that can only disconfirm the rule. In the example, one must turn over cards with odd numbers because the rule is disconfirmed if any of those cards have a vowel on the other side. In contrast, cards that can only confirm the rule are irrelevant. (One should not turn over cards with even numbers because the rule is technically not disconfirmed no matter whether there is a consonant or vowel on the other side.)

Chen and Cowan (in press) found performance on a working memory task that closely resembles confirmation bias. In one procedure, a spatial array of letters was presented on each trial, followed by a set of all of the letters at the bottom of the screen and a single location marked; the task was to select the correct letter for the marked location. In another procedure, the spatial array of letters was followed by a single letter from the array at the bottom of the screen and all of the locations marked; the task was to select the correct location for the presented letter. When working memory does not happen to contain the probed item, these procedures allow the use of disconfirming information. In the first task, for example, a participant might reason as follows: *The letters were K, R, Q, and L. I know the locations of only R and L and neither of them match the probed location. Therefore, I know that the answer must be K or Q and I will guess randomly between them*. That would be comparable to using disconfirming evidence. The pattern of data, however, did not appear to indicate that kind of process. Instead, participants answered correctly if they knew the probed item and otherwise guessed randomly among all of the other choices, without using the process of elimination. A mathematical model that assumed the latter process showed near-perfect convergence in capacity between the procedures described above and the usual change-detection procedure. If we instead assumed a mathematical model of performance in which disconfirming evidence was used through the process of elimination, there was no such convergence between the procedures.

So in reasoning and in working memory, processing tends to be inefficient, and it remains to be seen whether it can be meaningfully improved in terms of eliminating confirmation bias. Perhaps people with insufficient working memory or intelligence will always be stuck in such inefficient reasoning and there is nothing we can do. Arguing against that pessimistic view, however, is the recent finding (Stanovich, West, & Toplak, 2013) that the tendency to evaluate evidence more favorably when it agrees with one's own view occurs across the

board and is not correlated with intelligence, and presumably therefore not correlated with working memory either. One might be able to train individuals to make the best use of the working memory they have without worrying about increasing the basic capacity of working memory, either by training critical thinking skills (Halpern, 1989) or by instilling expertise (Eriksson et al., 2004).

Conclusion

Working memory is the retention of a small amount of information in a readily accessible form, which facilitates planning, comprehension, reasoning, and problem-solving. When we talk of working memory, we often include not only the memory itself, but also the executive control skills that are used to manage information in working memory and the cognitive processing of information. Theoretically, there is still uncertainty about the basic limitations on working memory: are they limitations on concurrent holding capacity, mnemonic processing speed, duration of retention of information before it decays, or just the same sorts of interference properties that apply to long-term memory? While these basic issues are debated and empirical investigations continue, there is much greater agreement about what results are obtained in particular test circumstances; the results of working memory studies seem rather replicable, but small differences in method produce large differences in results, so that one cannot assume that a particular working memory finding is highly generalizable.

For learning and education, it is important to take into account the basic principles of cognitive development and cognitive psychology, adjusting the materials to the working memory capabilities of the learner. We are not yet at a point at which every task can be analyzed in advance in order to predict which tasks are doable with a particular working memory capability. It is possible, though, to monitor performance and keep in mind that failure could be due to working memory limitations, adjusting the presentation accordingly. Keeping in mind the limitations of working memory of listeners and readers could easily help to improve one's lecturing and writing styles. I hope that awareness of working memory leads to a world in which we are all more tolerant of one another's inability to understand perfectly, are more humble and less arrogant, and are better able to communicate, educate one another, and reach common ground.

Acknowledgments

References

Andrews G, Halford GS. A cognitive complexity metric applied to cognitive development. Cognitive Psychology. 2002; 45:153–219. [PubMed: 12528901]

Atkinson, RC.; Shiffrin, RM. Human memory: A proposed system and its control processes. In: Spence, KW.; Spence, JT., editors. The psychology of learning and motivation: Advances in research and theory. Vol. Vol. 2. New York: Academic Press; 1968. p. 89-195.

Baddeley, AD. Working memory. Oxford: Clarendon Press; 1986. Oxford Psychology Series #11

Baddeley A. The episodic buffer: a new component of working memory? Trends in Cognitive Sciences. 2000; 4:417–423. [PubMed: 11058819]

Baddeley AD, Gathercole SE, Papagno C. The phonological loop as a language learning device. Psychological Review. 1998; 105:158–173. [PubMed: 9450375]

Barrouillet P, Gavens N, Vergauwe E, Gaillard V, Camos V. Working memory span development: A time-based resource-sharing model account. Developmental Psychology. 2009; 45:477–490. [PubMed: 19271832]

Baddeley, AD.; Hitch, G. Working memory. Bower, GH., editor. Vol. Vol. 8. New York: Academic Press; 1974. p. 47-89.The psychology of learning and motivation

Baddeley A, Papagno C, Vallar G. When long-term learning depends on short-term storage. Journal of Memory and Language. 1988; 27:586–595.

Barrouillet P, Portrat S, Camos V. On the law relating processing to storage in working memory. Psychological Review. 2011; 118:175–192. [PubMed: 21480738]

Bauer, PJ.; Fivush, R., editors. Handbook on the development of children's memory. Wiley-Blackwell; in press.

Bolton TL. The growth of memory in school children. American Journal of Psychology. 1892; 4:362–380.

Borst G, Niven E, Logie RH. Visual mental image generation does not overlap with visual short-term memory: A dual-task interference study. Memory & cognition. 2012; 40:360–372. [PubMed: 21989739]

Broadbent, DE. Perception and communication. New York: Pergamon Press; 1958.

Brown GDA, Neath I, Chater N. A temporal ratio model of memory. Psychological Review. 2007; 114:539–576. [PubMed: 17638496]

Burgess N, Hitch GJ. Memory for serial order: A network model of the phonological loop and its timing. Psychological Review. 1999; 106:551–581.

Camos V, Barrouillet P. Developmental change in working memory strategies: From passive maintenance to active refreshing. Developmental Psychology. 2011; 47:898–904. [PubMed: 21417518]

Camos V, Mora G, Oberauer K. Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. Memory & Cognition. 2011; 39:231–244. [PubMed: 21264630]

Case R, Kurland DM, Goldberg J. Operational efficiency and the growth of short-term memory span. Journal of Experimental Child Psychology. 1982; 33:386–404.

Chein J, Morrison A. Expanding the mind's workspace: training and transfer effects with a complex working memory span task. Psychonomic Bulletin & Review. 2010; 17:193–199. [PubMed: 20382919]

Chen Z, Cowan N. Core verbal working memory capacity: The limit in words retained without covert articulation. Quarterly Journal of Experimental Psychology. 2009; 62:1420–1429.

Chen Z, Cowan N. Working memory inefficiency: Minimal information is utilized in visual recognition tasks. Journal of Experimental Psychology: Learning, Memory, & Cognition. 2013; 39:1449–1462.

Chi, MTH. Knowledge structures and memory development. Siegler, R., editor. Hillsdale, NJ: Erlbaum; 1978. Children(s thinking: What develops?

Clark EV, Garnica OK. Is he coming or going? On the acquisition of deictic verbs. Journal of Verbal Learning and Verbal Behavior. 1974; 13:559–572.

Cocchini G, Logie RH, Della Sala S, MacPherson SE, Baddeley AD. Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. Memory & Cognition. 2002; 30:1086–1095. [PubMed: 12507373]

Conrad R. Acoustic confusion in immediate memory. British Journal of Psychology. 1964; 55:75–84.

Courage, ML.; Cowan, N., editors. The development of memory in infancy and childhood. Hove, U.K: Psychology Press; 2009.

Cowan N. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. Psychological Bulletin. 1988; 104:163–191. [PubMed: 3054993]

Cowan N. Verbal memory span and the timing of spoken recall. Journal of Memory and Language. 1992; 31:668–684.

Cowan, N. Attention and memory: An integrated framework. New York: Oxford University Press; 1995. Oxford Psychology Series, No. 26

Cowan, N. An embedded-processes model of working memory. In: Miyake, A.; Shah, P., editors. Models of Working Memory: Mechanisms of active maintenance and executive control. Cambridge, U.K: Cambridge University Press; 1999. p. 62-101.

Cowan N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences. 2001; 24:87–185. [PubMed: 11515286]

Cowan, N. Working memory capacity. Hove, East Sussex, UK: Psychology Press; 2005.

Cowan N. The magical mystery four: How is working memory capacity limited, and why? Current Directions in Psychological Science. 2010; 19:51–57. [PubMed: 20445769]

Cowan N. The focus of attention as observed in visual working memory tasks: Making sense of competing claims. Neuropsychologia. 2011; 49:1401–1406. [PubMed: 21277880]

Cowan N, AuBuchon AM, Gilchrist AL, Ricker TJ, Saults JS. Age differences in visual working memory capacity: Not based on encoding limitations. Developmental Science. 2011; 14:1066–1074. [PubMed: 21884322]

Cowan N, Donnell K, Saults JS. A list-length constraint on incidental item-to-item associations. Psychonomic Bulletin & Review. in press

Cowan N, Li D, Moffitt A, Becker TM, Martin EA, Saults JS, Christ SE. A neural region of abstract working memory. Journal of Cognitive Neuroscience. 2011; 23:2852–2863. [PubMed: 21261453]

Cowan N, Morey CC. How can dual-task working memory retention limits be investigated? Psychological Science. 2007; 18:686–688. [PubMed: 17680938]

Cowan N, Morey CC, AuBuchon AM, Zwilling CE, Gilchrist AL. Seven-year-olds allocate attention like adults unless working memory is overloaded. Developmental Science. 2010; 13:120–133. [PubMed: 20121868]

Cowan N, Nugent LD, Elliott EM, Ponomarev I, Saults JS. The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. Child Development. 1999; 70:1082–1097. [PubMed: 10546336]

Cowan N, Rouder JN, Blume CL, Saults JS. Models of verbal working memory capacity: What does it take to make them work? Psychological Review. 2012; 119:480–499. [PubMed: 22486726]

Cowan N, Saults JS, Winterowd C, Sherk M. Enhancement of 4-year-old children's memory span for phonologically similar and dissimilar word lists. Journal of Experimental Child Psychology. 1991; 51:30–52. [PubMed: 2010726]

Craik FIM, Watkins MJ. The role of rehearsal in short-term memory. Journal of Verbal Learning and Verbal Behavior. 1973; 12:599–607.

Daneman M, Carpenter PA. Individual differences in working memory and reading. Journal of Verbal Learning & Verbal Behavior. 1980; 19:450–466.

Diamond A, Lee K. Interventions shown to aid executive function development in children 4 to 12 years old. Science. 2011; 333:959–964. [PubMed: 21852486]

Ebbinghaus, H. Memory: A contribution to experimental psychology. Ruger, HA.; Bussenius, CE., translators. New York: Teachers College, Columbia University; 1885/1913. (Originally in German, *Ueber das gedächtnis: Untersuchen zur experimentellen psychologie*)

Engle RW, Tuholski SW, Laughlin JE, Conway ARA. Working memory, short term memory, and general fluid intelligence: A latent variable approach. Journal of Experimental Psychology: General. 1999; 128:309–331. [PubMed: 10513398]

Ericsson KA, Chase WG, Faloon S. Acquisition of a memory skill. Science. 1980; 208:1181–1182. [PubMed: 7375930]

Ericsson KA, Delaney PF, Weaver G, Mahadevan R. Uncovering the structure of a memorist's superior basic-memory capacity. Cognitive Psychology. 2004; 49:191–237. [PubMed: 15342260]

Ericsson KA, Kintsch W. Long-term working-memory. Psychological Review. 1995; 102:211–245. [PubMed: 7740089]

Farrell S, Lewandowsky S. An endogenous distributed model of ordering in serial recall. Psychonomic Bulletin & Review. 2002; 9:59–79. [PubMed: 12026954]

Flavell JH, Beach DH, Chinsky JM. Spontaneous verbal rehearsal in a memory task as a function of age. Child Development. 1966; 37:283–299. [PubMed: 5941895]

Fougnie D, Marois R. What limits working memory capacity? Evidence for modality-specific sources to the simultaneous storage of visual and auditory arrays. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2011; 37:1329–1341.

Garcia L, Nussbaum M, Preiss DD. Is the use of information and communication technology related to performance in working memory tasks? Evidence from seventh-grade students. Computers & Education. 2011; 57:2068–2076.

Garrity LI. An electromyographical study of subvocal speech and recall in preschool children. Developmental Psychology. 1975; 11:274–281.

Gathercole SE, Baddeley AD. Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. Journal of Memory and Language. 1989; 28:200–213.

Gathercole SE, Baddeley AD. Phonological memory deficits in language disordered children: Is there a causal connection? Journal of Memory and Language. 1990; 29:336–360.

Gathercole, SE.; Lamont, E.; Alloway, TP. Working memory in the classroom. In: Pickering, SJ., editor. Working memory and education. San Diego: Academic Press; 2006. p. 219-240.

Gathercole SE, Pickering SJ, Ambridge B, Wearing H. The structure of working memory from 4 to 15 years of age. Developmental Psychology. 2004; 40:177–190. [PubMed: 14979759]

Geary DC. An evolutionarily informed education science. Educational Psychologist. 2008; 43:179–195.

Gelman R. Accessing one-to-one correspondence: still another paper about conservation. British Journal of Psychology. 1982; 73:209–220.

Gilchrist AL, Cowan N, Naveh-Benjamin M. Investigating the childhood development of working memory using sentences: New evidence for the growth of chunk capacity. Journal of Experimental Child Psychology. 2009; 104:252–265. [PubMed: 19539305]

Halford GS, Cowan N, Andrews G. Separating cognitive capacity from knowledge: A new hypothesis. Trends in Cognitive Sciences. 2007; 11:236–242. [PubMed: 17475538]

Halford GS, Wilson WH, Phillips S. Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. Behavioral and Brain Sciences. 1998; 21:803–865. [PubMed: 10191879]

Hall, VC. Educational psychology from 1890 to 1920. In: Zimmerman, BJ.; Schunk, DH., editors. Educational psychology: A century of contributions. Mahwah, NJ: Erlbaum; 2003.

Halpern DF. Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacogitive monitoring. American Psychologist. 1998; 53:449–455. [PubMed: 9572008]

Halpern DF, Millis K, Graesser AC, Butler H, Forsyth C, Cai Z. Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. Thinking Skills And Creativity. 2012; 7:93–100.

Hebb, DO. Organization of behavior. New York: Wiley; 1949.

Henry LA. The effects of word length and phonemic similarity in young children's short-term memory. Quarterly Journal of Experimental Psychology. 1991; 43A:35–52.

Henson RNA. Short-term memory for serial order: The start-end model. Cognitive Psychology. 1998; 36:73–137. [PubMed: 9721198]

Hulme C, Tordoff V. Working memory development: The effects of speech rate, word length, and acoustic similarity on serial recall. Journal of Experimental Child Psychology. 1989; 47:72–87.

Jaeggi, SM.; Buschkuehl, M. Training working memory. In: Alloway, TP.; Alloway, RG., editors. Working memory: The connected intelligence. NY: Psychology Press; 2013.

James, W. The principles of psychology. NY: Henry Holt; 1890.

Jarrold C, Citroën R. Reevaluating key evidence for the development of rehearsal: Phonological similarity effects in children are subject to proportional scaling artifacts. Developmental Psychology. 2013; 49:837–847. [PubMed: 22662766]

Jevons WS. The power of numerical discrimination. Nature. 1871; 3:281–282.

Jolicoeur P, Dell'Acqua R. The demonstration of short-term consolidation. Cognitive Psychology. 1998; 36:138–202. [PubMed: 9721199]

Kail, R. The development of memory in children. 3rd edition. NY: W.H.Freeman; 1990.

Kail R, Salthouse TA. Processing speed as a mental capacity. Acta Psychologica. 1994; 86:199–255. [PubMed: 7976467]

Kane MJ, Brown LH, McVay JC, Silvia PJ, Myin-Germeys I, Kwapil TR. For whom the mind wanders, and when: An experience-sampling study of working memory and executive control in daily life. Psychological Science. 2007; 18:614–621. [PubMed: 17614870]

Kane MJ, Engle RW. Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. Journal of Experimental Psychology: General. 2003; 132:47–70. [PubMed: 12656297]

Kane MJ, Hambrick DZ, Tuholski SW, Wilhelm O, Payne TW, Engle RW. The generality of workingmemory capacity: A latentvariable approach to verbal and visuospatial memory span and reasoning. Journal of Experimental Psychology: General. 2004; 133:189–217. [PubMed: 15149250]

Klingberg T. Training and plasticity of working memory. Trends in Cognitive Sciences. 2010; 14:317–324. [PubMed: 20630350]

Kyllonen PC, Christal RE. Reasoning ability is (little more than) working-memory capacity?! Intelligence. 1990; 14:389–433.

Kyndt E, Cascallar E, Dochy F. Individual Differences in Working Memory Capacity and Attention, and Their Relationship with Students' Approaches to Learning. Higher Education: The International Journal Of Higher Education And Educational Planning. 2012; 64:285–297.

Lewandowsky S, Duncan M, Brown GDA. Time does not cause forgetting in short-term serial recall. Psychonomic Bulletin & Review. 2004; 11:771–790. [PubMed: 15732687]

Locke, J. An essay concerning human understanding. London: Thomas Bassett; 1690.

Logie, RH. The seven ages of working memory. In: Richardson, JTE.; Engle, RW.; Hasher, L.; Logie, RH.; Stoltzfus, ER.; Zacks, RT., editors. Working memory and human cognition. New York: Oxford University Press; 1996. p. 31-65.

Logie, RH.; Van Der Meulen, M. Fragmenting and integrating visuospatial working memory. In: Brockmole, JR., editor. Representing the visual world in memory. Hove, U.K: Psychology Press; 2009. p. 1-32.

Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. Nature. 1997; 390:279–281. [PubMed: 9384378]

Mandler G, Shebo BJ. Subitizing: An analysis of its component processes Journal of Experimental Psychology. General. 1982; 111:1–22.

Melby-Lervåg M, Hulme C. Is working memory training effective? A meta-analytic review. Developmental Psychology. 2013; 49:270–291. [PubMed: 22612437]

Miller GA. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review. 1956; 63:81–97. [PubMed: 13310704]

Miller, GA.; Galanter, E.; Pribram, KH. Plans and the structure of behavior. New York: Holt, Rinehart and Winston, Inc; 1960.

Morey CC, Bieler M. Visual short-term memory always requires attention. Psychonomic Bulletin & Review. 2013; 20:163–170. [PubMed: 23002024]

Murdock BB. A distributed memory model for serial-order information. Psychological Review. 1983; 90:316–338.

Murdock BB, Walker KD. Modality effects in free recall. Journal of Verbal Leaning and Verbal Behavior. 1969; 8:665–676.

Nelson KJ. Variations in children's concepts by age and category. Child Development. 1974; 45:577–584. [PubMed: 4143813]

Norman DA. Toward a theory of memory and attention. Psychological Review. 1968; 75(6):522–536.

Oberauer K, Lewandowsky S. Forgetting in immediate serial recall: decay, temporal distinctiveness, or interference? Psychological Review. 2008; 115:544–576. [PubMed: 18729591]

Oberauer K, Lewandowsky S. Modeling working memory: A computational implementation of the time-based resource-sharing theory. Psychonomic Bulletin & Review. 2011; 18:10–45. [PubMed: 21327362]

Oberauer K, Lewandowsky S, Farrell S, Jarrold C, Greaves M. Modeling working memory: An interference model of complex span. Psychonomic Bulletin & Review. 2012; 19:779–819. [PubMed: 22715024]

Ornstein, PA.; Naus, MJ. Rehearsal processes in children's memory. In: Ornstein, PA., editor. Memory development in children. Hillsdale, NJ: Erlbaum; 1978. p. 69-99.

Paas F, Sweller J. 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. 2012; 24:27–45.

Pascual-Leone, J.; Johnson, J. A developmental theory of mental attention: Its applications to measurement and task analysis. In: Barrouillet, P.; Gaillard, V., editors. Cognitive development and working Memory: From neoPiagetian to cognitive approaches. Hove, UK: Psychology Press; 2011. p. 13-46.

Pascual-Leone J, Smith J. The encoding and decoding of symbols by children: A new experimental paradigm and a neo-Piagetian model. Journal of Experimental Child Psychology. 1969; 8:328–355.

Penney CG. Modality effects and the structure of short-term verbal memory. Memory & Cognition. 1989; 17:398–422. [PubMed: 2668697]

Pickering, SJ. Working memory and education. San Diego: Academic Press; 2006.

Redick TS, Shipstead Z, Harrison TL, Hicks KL, Fried DE, Hambrick DZ, Kane MJ, Engle RW. No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. Journal of Experimental Psychology: General. 2013; 142:359–379. [PubMed: 22708717]

Ricker TJ, Cowan N. Loss of visual working memory within seconds: The combined use of refreshable and non-refreshable features. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2010; 36:1355–1368.

Ricker TJ, Cowan N. Differences between presentation methods in working memory procedures: A matter of working memory consolidation. Journal of Experimental Psychology: Learning, Memory, and Cognition. in press

Rouder JN, Morey RD, Cowan N, Zwilling CE, Morey CC, Pratte MS. An assessment of fixed-capacity models of visual working memory. Proceedings of the National Academy of Sciences (PNAS). 2008; 105:5975–5979.

Sabol TJ, Pianta RC. Patterns of school readiness forecast achievement and socioemotional development at the end of elementary school. Child Development. 2012; 83:282–299. [PubMed: 22103310]

Saltz E, Soller E, Sigel IE. The development of natural language concepts. Child Development. 1972; 43:1191–1202. 1972.

Schüler A, Scheiter K, Genuchten E. The Role of Working Memory in Multimedia Instruction: Is Working Memory Working During Learning from Text and Pictures? Educational Psychology Review. 2011; 23:389–411.

Shim W, Walczak K. The Impact of Faculty Teaching Practices on the Development of Students' Critical Thinking Skills. International Journal Of Teaching And Learning In Higher Education. 2012; 24:16–30.

Shipstead Z, Redick TS, Engle RW. Is working memory training effective? Psychological Bulletin. 2012; 138:628–654. [PubMed: 22409508]

Slevc L. Saying What's on Your Mind: Working Memory Effects on Sentence Production. Journal Of Experimental Psychology: Learning, Memory, And Cognition. 2011; 37:1503–1514.

Sokolov, EN. Perception and the conditioned reflex. NY: Pergamon Press; 1963.

Stanovich KE, West RF, Toplak ME. Myside Bias, Rational Thinking, and Intelligence. Current Directions in Psychological Science. 2013; 22:259–264.

Süβ HM, Oberauer K, Wittmann WW, Wilhelm O, Schulze R. Working-memory capacity explains reasoning ability—and a little bit more. Intelligence. 2002; 30:261–288.

Sweller J. Cognitive load theory. Psychology of Learning and Motivation. 2011; 55:37–76.

Sweller J, van Merrienboer JJG, Paas FGWC. Cognitive architecture and instructional design. Educational Psychology Review. 1998; 10:251–296.

Tan L, Ward G. A recency-based account of the primacy effect in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2000; 26:1589–1625.

Todd JJ, Marois R. Capacity limit of visual short-term memory in human posterior parietal cortex. Nature. 2004; 428:751–754. [PubMed: 15085133]

Treisman AM. Contextual cues in selective listening. Quarterly Journal of Experimental Psychology. 1960; 12:242–248.

Tricot A, Sweller J. Domain-specific knowledge and why teaching generic skills does not work. Educational Psychology Review. in press

Van Der Meulen M, Logie RH, Sala SD. Selective interference with image retention and generation: Evidence for the workspace model. The Quarterly Journal of Experimental Psychology. 2009; 62:1568–1580. [PubMed: 19096989]

Vergauwe E, Barrouillet P, Camos V. Do mental processes share a domain-general resource? Psychological Science. 2010; 21:384–390. [PubMed: 20424075]

Vogel EK, McCollough AW, Machizawa MG. Neural measures reveal individual differences in controlling access to working memory. Nature. 2005; 438:500–503. [PubMed: 16306992]

Wason PC, Shapiro D. Natural and contrived experience in a reasoning problem. Quarterly Journal of Experimental Psychology. 1971; 23:63–71.

Wilding J. Over the top: Are there exceptions to the basic capacity limit? Behavioral and Brain Sciences. 2001; 24:152–153.

Wilhelm, O.; Engle, RW., editors. Handbook of understanding and measuring intelligence. London: Sage; 2005.

Wundt, W. Lectures on human and animal psychology. Creighton, JE.; Titchener, EB., editors. Bristol, U.K: Thoemmes Press; 1894/1998. Translated from the second German

Xu Y, Chun MM. Dissociable neural mechanisms supporting visual short-term memory for objects. Nature. 2006; 440:91–95. [PubMed: 16382240]

Zhang W, Luck SJ. Sudden death and gradual decay in visual working memory. Psychological Science. 2009; 20:423–428. [PubMed: 19320861]
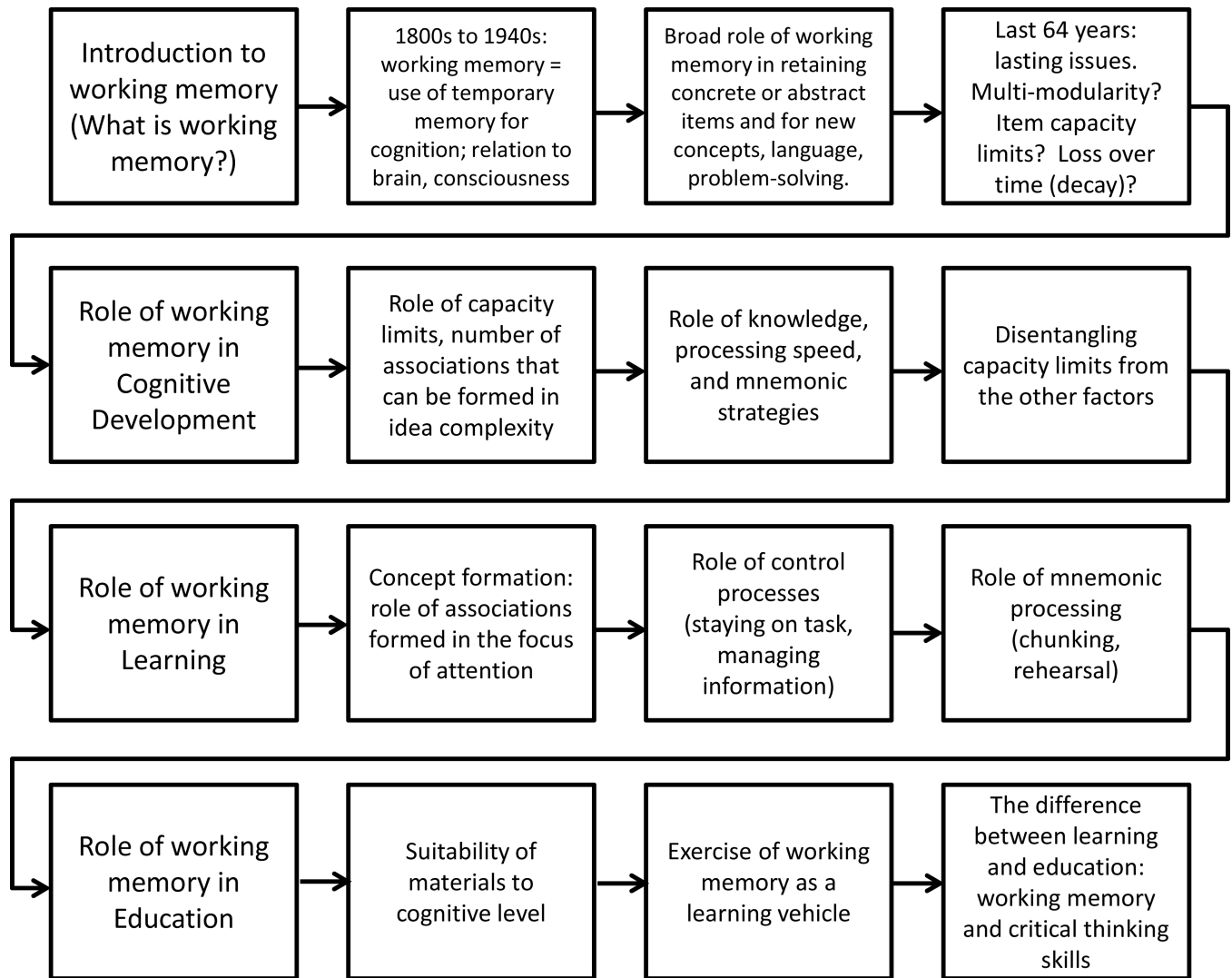
Introduction to working memory (What is working memory?)	→	1800s to 1940s: working memory = use of temporary memory for cognition; relation to brain, consciousness	→	Broad role of working memory in retaining concrete or abstract items and for new concepts, language, problem-solving.	→	Last 64 years: lasting issues. Multi-modularity? Item capacity limits? Loss over time (decay)?
Role of working memory in Cognitive Development	→	Role of capacity limits, number of associations that can be formed in idea complexity	→	Role of knowledge, processing speed, and mnemonic strategies	→	Disentangling capacity limits from the other factors
Role of working memory in Learning	→	Concept formation: role of associations formed in the focus of attention	→	Role of control processes (staying on task, managing information)	→	Role of mnemonic processing (chunking, rehearsal)
Role of working memory in Education	→	Suitability of materials to cognitive level	→	Exercise of working memory as a learning vehicle	→	The difference between learning and education: working memory and critical thinking skills

Figure 1.
Schematic diagram of the arguments in the present article.