

WORKING MEMORY MODELS

In their attempts to better understand the workings of the mind, psychologists develop explanatory models known as *constructs*. A hypothetical construct is inferred from data because it is not directly observable. For example, *intelligence* is a well-known and long-debated construct that cannot be directly observed or measured. This book is about *working memory* (WM), one of the most influential psychological constructs of the past 40 years. The behaviors associated with WM are measurable and real. However, the underlying construct associated with these behaviors remains hypothetical. Its exact nature, functioning, neurological structure, and even its name are still open to debate and refinement.

WM is the cognitive ability to briefly hold, maintain, or store information while processing the same or other information. Simply put, brief storage plus simultaneous cognitive processing equals WM. The brief storage aspect is commonly referred to as *short-term memory* (STM). Thus, the construct of WM includes STM, with WM having a supervisory role over the STM component (Baddeley, 1986). The supervisory role is just one of WM's executive functions. WM is complex; it has both cognitive and metacognitive dimensions (Dehn, 2014a).

What makes WM so interesting and so influential is that it is very limited in humans, and these limitations have significant consequences for all sorts of human endeavors. Without keeping information refreshed

DON'T FORGET

Working memory occurs whenever there is concurrent temporary storage and processing of information. Short-term memory provides the storage function. Thus, short-term memory is embedded within working memory. In this book, the term "working memory" includes short-term memory.

in WM, it will be retained only for a few seconds. In the typical adult, only four to seven pieces of information can be maintained in WM during cognitive processing (Cowan, 2001).

Psychologists were measuring WM long before the construct was even proposed. The *digit span* test goes back more than 100 years. This test includes *digits backward*, which requires the examinee to reverse the sequence of orally presented digits. Digits backward is now recognized as a robust measure of WM. Prior to the 1990s, the widely used *Wechsler Intelligence Scale for Children* incorporated the digit span subtest into a composite score it called *Freedom From Distractibility*, a label that describes one of WM's key functions but gave psychologists little understanding of what they were actually measuring.

CAUTION

Despite working memory's wide-ranging influence, the definition of working memory should remain narrow. For example, reasoning and working memory are not the same thing, even though reasoning heavily depends on working memory capacity.

When defining WM and discussing the roles that it plays in cognitive functioning, it is important to consider how WM is typically measured. The usual task requires the maintenance of oral or visual stimuli while processing those stimuli in some manner. What is actually measured is not the processing but the number of sequential items (a span) that is retained. All of the empirical data on

how WM is related with cognitive abilities, academic learning, and daily functioning is based on such traditional span measures of WM. Consequently, the definition and application of the WM construct should not go beyond how it is measured. For example, WM should not be equated with intelligence, general executive functioning, or all conscious mental activity. When the definition goes beyond the measurement of the construct, then WM becomes too inclusive and less meaningful. Also, very broad applications of the construct create false impressions that WM training should lead to improved functioning in psychological processes that may not really be WM.

WORKING MEMORY'S INFLUENCE

Nearly all cognitive and metacognitive functions are closely interrelated with WM. For example, language expression, processing speed, reasoning, phonological processing, attentional control, and executive functions have high correlations with WM (see Chapter 2). Furthermore, nearly all aspects of learning, especially academic learning, depend on adequate levels of WM (see Chapter 4).

Finally, performance and application of skills, as well as cognitively challenging daily activities, depend on WM. A short list of activities that are influenced by WM capacity includes:

- Keeping up with the flow of a conversation and remembering what one was going to say.
- Noticing errors that are contained in a written sentence one just produced.
- Keeping track of one's place while counting.
- Being able to take detailed notes while listening at the same time.
- Remembering multistep directions that were just presented or read.
- Completing a task in a time-efficient manner.
- Coping with distractions while thinking.
- Comprehending what is being said or read.
- Remembering what one was going to do next.
- Keeping track of subproducts while doing mental arithmetic.
- Being able to switch between mental tasks.
- Being able to reason, such as comparing and contrasting two concepts.
- Integrating visual and auditory information.
- Efficiently memorizing information.
- Consciously retrieving a name or word that does not come immediately.

Obviously, a normal WM ability is essential for all kinds of cognitive, learning, and daily activities (Engle, 2002). Consequently, unusual shortcomings or deficits in WM can lead to all kinds of problems. Such problems include forgetfulness, inattentiveness, difficulty following directions, difficulty completing tasks, difficulty communicating, and various types of learning disorders.

BADDELEY'S WORKING MEMORY MODEL

The predominant model of WM was originally proposed by Baddeley and Hitch in 1974 and later expanded by Baddeley (1986, 2000). The Baddeley model of WM is the theoretical basis of the majority of research on WM. This multicomponent model has been validated through neuropsychological research and has been operationalized in measurement instruments. Baddeley defines *working memory* as “a system for the temporary holding and manipulation of information during the performance of a range of cognitive tasks such as comprehension, learning, and reasoning” (1986, p. 34). The original multifaceted model was made up of three components: a phonological loop, a visuospatial sketchpad, and a central executive. In 2000, Baddeley added another component—the episodic buffer (see Rapid Reference 1.1). Baddeley's model is hierarchical, with the central executive as the top-level, domain-free facet that controls all of the subcomponents.

≡ Rapid Reference 1.1 Baddeley's Working Memory Model

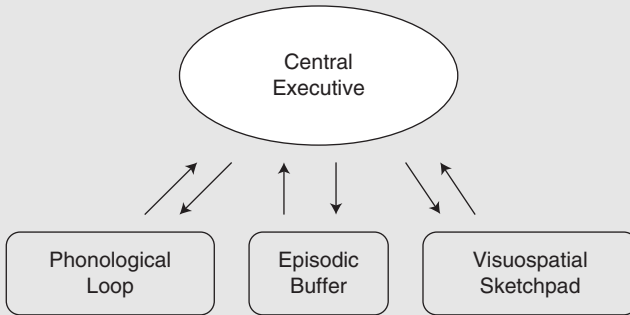


Figure 1.1 Baddeley's Working Memory Model

The Phonological Loop

What Baddeley refers to as the *phonological loop* is also known as auditory, phonological, or verbal short-term memory. (In this text, this aspect of WM will be called *phonological short-term memory*.) The phonological loop is a limited capacity component that briefly stores verbal information in phonological form. Baddeley (1986, 2003a) divides the phonological loop into passive storage and subvocal, articulatory rehearsal. The number of unconnected verbal items (such as words from a list) that can be retained in the phonological loop depends on the time it takes to articulate them (Baddeley, 2003a). Individuals can recall only a sequential span that they can articulate (aloud or subvocally) within 2 seconds (Ellis & Hennesley, 1980; Hulme & Mackenzie, 1992). For instance, if an individual's speech rate is two monosyllable words per second, his memory span will be about four monosyllable words. Thus, auditory STM span varies according to the length of the words and the individual's speech rate. Individuals with faster articulation rates can maintain more items than individuals who are slow articulators (Hulme & Mackenzie, 1992). Also, more monosyllable words can be retained than multisyllable words. For adults, normal phonological loop span is approximately seven monosyllable units.

Despite the strong evidence that word length and articulatory rehearsal speed determine auditory STM span, other influences also affect memory span. One

influence is prior knowledge. Meaningful phonological information may activate relevant long-term memory (LTM) representations, which may then facilitate immediate recall from short-term storage. For instance, the average adult has a longer span for meaningful words than for pseudo-words. The degree of chunking or grouping of items into larger units also affects span. For example, the separate digits “five” and “eight” can be chunked as “fifty-eight.” Also, individuals can remember sentences that take several seconds to articulate because the sentences can be chunked into meaningful phrases or ideas.

Subvocal rehearsal seems to largely determine verbal span because whenever individuals are prevented from rehearsing, performance is markedly impaired. The typical interference task prevents rehearsal by requiring the participant to engage in an unrelated attention-demanding task, such as counting. The impact of disrupting phonological short-term rehearsal provides evidence of the importance of rehearsal to the short-term retention of information. Without rehearsal, less information will be retained, and the retention interval will last only a few seconds (Henry, 2001).

The Visuospatial Sketchpad

What Baddeley refers to as the *visuospatial sketchpad* is also known as *visual-spatial short-term memory* (the label that will be used in this text). The visuospatial sketchpad is responsible for the immediate storage of visual and spatial information, such as the color of objects and their location. Like the phonological component, it consists of passive temporary storage and mostly automated rehearsal. Decay in the visuospatial sketchpad seems to be as rapid as phonological decay, taking place within a matter of seconds. Rehearsal of the information involves eye movement, manipulation of the image, or some type of visual mnemonic (Baddeley, 1986).

Neurologically, visual-spatial short-term storage has two dimensions: visual and spatial (Pickering, Gathercole, Hall, & Lloyd, 2001). The visual aspect is responsible for the storage of static visual information (e.g., shape, color, and size). The spatial dimension is responsible for the storage of dynamic information (e.g., location, motion, and direction). Visual-spatial short-term storage capacity is typically three to seven items for a matter of seconds.

Complex or abstract stimuli are more difficult to retain than simple or common stimuli that can be named. For example, blocks displayed in a matrix are easier to recall than a random display, and abstract figures are more difficult to recall than drawings of common objects. These findings indicate that structured visual-spatial information consumes less short-term storage capacity than unstructured. The fact that visual-spatial span is better for familiar material suggests that long-term

memory representations are facilitating short-term visual-spatial memory, much like recognizable words extend phonological memory span.

Better recall for familiar images may also be accomplished by the recoding of visual-spatial information into verbal information, which is more likely to occur when images are recognizable and can be named. Once recoded, verbal rehearsal maintains the information that originally was visual-spatial (Richardson, 1996). By 10 years of age, most children verbally recode visual-spatial stimuli.

In Baddeley's model, the generation, manipulation, and maintenance of visual images also appear to involve the visuospatial sketchpad (De Beni, Pazzaglia, Meneghetti, & Mondoloni, 2007). Maintenance and manipulation of visual images are highly demanding, requiring more than the visuospatial sketchpad itself. Thus, WM's central executive must become involved whenever internally generated visual images are being consciously manipulated.

The Episodic Buffer

Baddeley's (2000) *episodic buffer* refers to the interaction between WM and LTM. This interaction takes new information currently being held and processed in WM and incorporates it with already existing LTM representations. For example, the episodic component combines visual and verbal codes and links them to multidimensional representations in LTM. The episodic buffer is involved in learning because this is where encoding into LTM takes place (Pickering & Gathercole, 2004). It is also involved in conscious efforts to retrieve desired information from LTM storage.

The episodic buffer can account for temporary storage of large amounts of information that exceed the capacities of the phonological and visuospatial storage systems (Baddeley, 2003a). This perspective is best explained by Cowan's (2005) *embedded process model*. Cowan proposes that there is a pool of recently activated LTM items. WM interacts with these activated pieces of information, quickly switching the focus of attention from one to another as the processing task demands change. Thus, the episodic buffer accounts for how individuals can handle more information than would be indicated by measures of WM span. Unfortunately, a standardized method of measuring the capacity of the episodic buffer has not yet been devised. Consequently, further discussion of the episodic buffer in this book will be limited.

The Central Executive

In Baddeley's model, the *central executive* is responsible for managing the three subcomponents, and it also regulates and coordinates all of the cognitive

subprocesses involved in WM performance. The central executive is involved whenever an individual must simultaneously store and process information (Tronsky, 2005). For example, the central executive is responsible for managing dual-task situations, which typically involve processing information while trying to retain the same or different information. As described by Baddeley (1996b), the central executive is multimodal and does not have its own temporary storage, relying on the phonological loop and visual-spatial sketchpad for storage. In addition, it has limited capacity for the processes it conducts. Overall, the primary role of the central executive is to coordinate information from a number of different sources and manage performance on separate, simultaneous tasks (Baddeley, 1996b). For example, the central executive will be involved whenever visual and verbal information needs to be integrated.

Over the years, Baddeley (1986, 1996b, 2003b, 2006) has described several core central executive functions, including (a) focusing attention on relevant information while inhibiting the irrelevant information; (b) switching between concurrent cognitive activities; (c) applying strategies, such as conscious rehearsal; (d) allocating limited resources to other parts of the WM system; and (e) retrieving, holding, and manipulating temporarily activated information from LTM.

OTHER MODELS OF WORKING MEMORY

Although Baddeley's model is the most popular and has the most empirical and neurological support, there are other conceptualizations of how WM functions. Most of these have some empirical support as well. They also account for WM performance in ways that further enhance our understanding of WM. The other models mainly differ from Baddeley's in that they emphasize processing instead of storage, the control of attention, and the interactions between WM and LTM.

Daneman and Carpenter's Processing Efficiency Model

Daneman and Carpenter (1980) emphasize the processing dimension of WM, arguing that what appears to be smaller storage capacity may actually be the result of inefficient processing. They contend that complex mental operations utilize WM resources and the more efficient the mental processing, the more resources are available for short-term storage. Because processing efficiency varies by task, WM performance varies, depending on the task at the moment. For example, an expert in chess has better WM-related performance during a chess match than a novice does. However, the chess expert will not excel when a nonchess task is used to assess WM. Daneman and Carpenter believe that individuals do not vary

as much in available storage capacity as they do in processing efficiency. They also believe that storage and processing capacity remain relatively constant during development. Age-related improvements in span result mainly from increased operational efficiency. Although this model views WM as including both storage and processing functions, the model reduces the need for modality-specific storage buffers (Just & Carpenter, 1992). For Daneman and Carpenter, WM essentially corresponds to the central executive in Baddeley's theory. From their perspective, performance on complex span tasks is due primarily to central executive processing efficiency, not storage capacity.

Kane and Engle's Executive Attention Model

Kane and Engle (Engle 1996, 2002; Kane, Conway, Bleckley, & Engle, 2001) portray WM as an executive attention function that is distinguishable from STM. Kane and Engle make the case that WM capacity is not about short-term span but rather about the ability to control attention in order to maintain information in an active, quickly retrievable state. They define *executive attention* as "an executive control capability; that is, an ability to effectively maintain stimulus, goal, or context information in an active, easily accessible state in the face of interference, to effectively inhibit goal-irrelevant stimuli or responses, or both" (Kane et al., 2001, p. 180). Executive attention not only allows switching between competing tasks but also maintains desired information by suppressing and inhibiting unwanted, irrelevant information. Therefore, the capacity of WM is a function of how well executive processes can focus attention on the relevant content and goals, not on the length of the retention interval or how much short-term storage is available.

Evidence for their model comes from studies in which high-memory-span participants demonstrate better attentional control than low-span subjects. High-span individuals are more adept at resisting interference than low-span subjects (Kane et al., 2001). Their ability to inhibit interference allows them to retain and process more information. Most of the interference is internally generated, often caused by associating current information with earlier information that is no longer relevant. Thus, individuals with a high WM span may not necessarily have a greater short-term storage capacity than those with a low span. Rather, WM span is constrained by the executive capacity to control attention and resist interference (Hester & Garavan, 2005).

Kane and Engle (2000) also emphasize the role of WM in retrieving and actively maintaining information from LTM. They believe WM is responsible for cue-dependent, focused searching that has a high probability of leading to correct

recall. Furthermore, this cue-dependent process applies to retrieval of information just recently lost from short-term storage because of the removal of attention, extended time intervals, or distractions. Such information has often transferred to the pool of recently activated LTM items.

According to Kane and Engle (2000), low-WM-capacity individuals have more difficulty selecting and using correct cues to guide the LTM search process, resulting in too many irrelevant representations being retrieved and ultimately in failure to retrieve the sought-after information. Thus, individual differences in WM capacity are also related to individual differences in the ability to engage in a controlled, strategic search of LTM (Unsworth & Engle, 2007).

Kane and Engle emphasize a strong connection between WM and LTM. They view WM as a subset of recently activated LTM units (Unsworth & Engle, 2007). By continually focusing attention, WM maintains a few representations (typically about four) for ongoing processing. As attentional resources increase with age, more LTM structures can be activated concurrently.

Kane and Engle (2000) have also investigated the relationships WM has with higher-level cognitive functions. According to their theory, controlled attention is what binds all of the cognitive processes and functions, such as fluid reasoning, together. In contrast, the STM components are not significantly related to higher-level cognition.

In summary, Kane and colleagues are proposing that WM consists of domain-general controlled attention, which is mainly applied to retrieving and maintaining activation of LTM representations. Individual differences in WM reflect the degree to which distracters can be inhibited and relevant information can be actively maintained as the focus of attention (Kane et al., 2001). The theory makes inhibitory control the primary determinant of working memory capacity.

DON'T FORGET

Most models of working memory emphasize the close interaction between working memory and long-term memory. Working memory is constantly utilizing items from a recently activated pool of long-term memory representations. This activated pool can contain more than 20 items at a time (Anderson, 1983), thereby expanding the amount of information that working memory can handle.

Cowan's Embedded-Process Model

Cowan (2005) is an American psychologist who has greatly expanded the construct of WM, altered the view of WM capacity, and closely linked WM with LTM. His model emphasizes focus of attention, levels of activation, and

expertise as essential properties of WM. Cowan's theory embeds WM within LTM while still recognizing WM and STM as separable from LTM. Essentially, Cowan believes that WM refers to information in LTM that is activated above some threshold.

Cowan's model mainly distinguishes between the activated part of LTM and the focus of attention (see Rapid Reference 1.2). Only the focus of attention has limited capacity, typically a few highly activated elements at a time. The larger pool of activated LTM items is not capacity-limited, but items can be lost through decay or interference (Oberauer, 2002). Items in the activated pool quickly move in and out of the focus of attention, depending on what is needed at the moment.

The focus of attention replaces the multiple separate storage buffers and the central executive of Baddeley's model. Cowan posits that a limited focus of attention restricts WM retention and processing, not storage capacity. In adults with normal WM capacity, the focus of attention can handle about four chunks of activated information at a time. Studies of retrieval speed (McElree, 1998) provide support for Cowan's model by finding that items expected to be in the focus of attention are retrieved more quickly than recently activated items that are no longer the focus of attention.

Rapid Reference 1.2 Cowan's Embedded-Process Model

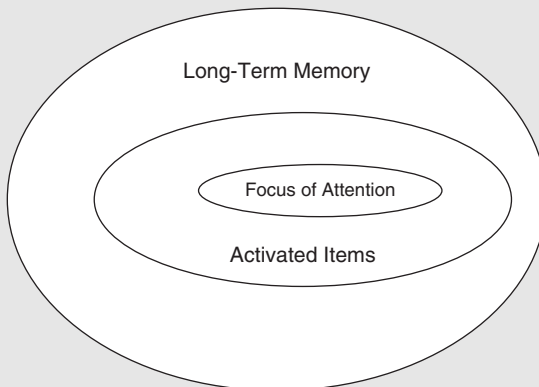


Figure 1.2 Cowan's Embedded-Process Model

Ericsson and Kintsch's Long-Term Working Memory Model

Given the close connection between WM and LTM, it is not surprising that there are advocates (Ericsson & Kintsch, 1995) for a *long-term working memory*. According to this view, WM is not structurally distinct from LTM. Essentially, WM is the skillful utilization of information stored in LTM. Although WM may not be separable from LTM in this model, it still performs the same functions, such as processing select sensory input and encoding new information into long-term storage.

The notion of long-term WM changes the perspective on storage capacity. Instead of how many chunks can be held in short-term storage, capacity entails how many long-term representations can be in a highly active state at any one time (Richardson, 1996). As suggested by Cowan (2005), the typical individual can hold and manipulate about four pieces of information concurrently. Similar to decay in STM models, activated portions of LTM must quickly return to an inactive state so that there is room for other long-term representations as they become activated. This perspective opens the door to the possibility that much of what is immediately retrieved is actually being retrieved from long-term, not short-term, storage. This leads to several educational implications, among them the benefits of long-term mnemonics on WM functioning.

Ericsson and Kintsch suggest that the skillful use of information held in LTM depends on expertise and the use of mnemonics, both of which enable individuals to use LTM as an efficient extension of WM. Extended WM seems to depend mainly on grouping items into chunks and then associating the chunks with familiar patterns, such as schemas, already stored in LTM. The process of reading comprehension is consistent with Ericsson and Kintsch's model. Comprehension over extended parts of text would not be possible without LTM involvement. As the reader progresses through text, a representation is constructed in LTM. This representation is continually expanded to integrate new information from the text, with relevant parts remaining accessible during reading. Ericsson and Kintsch view the accessible portions of this structure as an extended WM. Their argument seems plausible, given that text comprehension increases dramatically from childhood to adulthood without concomitant increases in short-term and WM capacity.

The Relationship Between Working Memory and Long-Term Memory

Most of the long-term WM theorists agree that WM is a separate cognitive process, even if it might be embedded within LTM. In contrast, advocates for a separate WM view LTM and WM as structurally separate memory systems, each

with its own storage area. This viewpoint is consistent with neuroscience evidence. For example, when individuals work at WM tasks, it is the dorsolateral prefrontal cortex and related STM storage areas that are the most active, rather than the hippocampus, which is an LTM structure (see Chapter 3 for neuroanatomy details). Also, amnesic cases illustrate (Corkin, 1984) that STM and WM can function adequately while LTM is impaired. However, there is no doubt that WM and LTM are highly interrelated. Thus, models of WM that focus only on the WM-STM connection may be missing valuable insights into the functioning and dysfunctioning of human memory systems.

THE CONTROVERSY REGARDING WORKING MEMORY CAPACITY

Most models represent WM as a unitary system where processing and storage compete for a limited, common pool of resources. This shared-resources position is known as the *general capacity hypothesis* (Engle, Cantor, & Carullo, 1992). On a moment-to-moment, flexible basis, resources are shared between the STM and LTM storage components and the executive component. Unless the individual consciously prioritizes storage, processing demands receive priority. When the processing demands of the task are high, such as trying to solve a complex mental arithmetic problem, WM capacity cannot meet demand. The result is information loss, an inability to complete the task, or, at the very least, slower processing (Engle et al., 1992). In typical cognitive activities, the difficulty of the processing task is inversely related to memory span.

Humans are capable of simultaneously storing and processing information. In order to do so, they must continually and rapidly switch back and forth between storage and processing (Barrouillet, Gavens, Vergauwe, Gaillard, & Camos, 2009). During the storage phase they are rehearsing the material to prevent decay, and during the processing phase at least some of the items in storage are neglected. Accordingly, span reduction results from limited, disrupted, infrequent, or too-late rehearsal opportunities. Thus, the amount of information an individual can simultaneously maintain and process partly depends on the efficiency of switching. However, attention-demanding processes reduce the proportion of time that can be allocated to rehearsal (Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007). On the other hand, efficient processing releases additional resources for maintaining storage, thereby increasing memory span. Therefore, improvements in WM performance may be due to increased processing efficiency, rather than increases in storage capacity per se.

Conscious and systematic switching is challenging. Switching is also a process that draws from overall WM capacity. Thus, more efficient and automated

switching detracts less from storage and the task at hand. Most children are unable to switch back and forth prior to the age of 7 (Barrouillet et al., 2009). At early ages, children focus only on the processing task, at the expense of maintaining information in STM. Switching efficiency apparently develops gradually. The rate of switching so that rehearsal can occur is twice as high for 14-year-old children as it is for 8-year-old children (Barrouillet et al., 2009). When children are developmentally ready, they can be taught how to switch. WM exercises that incorporate switching training are discussed in Chapter 7. In contrast to the position that WM has limited resources that are shared between storage and processing, some cognitive psychologists (Halford, Wilson, & Phillips, 2001) postulate that there are separate capacity limits for storage (STM components) and processing (executive WM). This claim, known as the *separate resources hypothesis*, is in accord with the beliefs that storage and processing demands are quite different. From this viewpoint, the capacity of the central executive determines the rate and extent of information processing, whereas STM span reflects the storage capacity of the phonological loop or visuospatial sketchpad (Baddeley, 1990). Short-term storage components each have their own capacity limitations, which are distinct from executive WM capacity.

There is empirical evidence supporting the separate resources hypothesis. As expressed by Towse, Hitch, and Hutton (1998), “Storage is independent of concurrent processing load, and processing performance is independent of concurrent storage load. The relationship between processing and storage arises because the time spent in processing affects the amount of forgetting that accrues” (p. 219). Also, there are numerous examples (reviewed by Oberauer, 2002) of short-term retention being unimpaired by concurrent secondary processing tasks. Moreover, the amount of information retained in storage seems unaffected by the degree of processing. Even in demanding dual-task experimental designs, participants typically perform well on both storage and processing (Seigneuric, Ehrlich, Oakhill, & Yuill, 2000). In instances where there is a decrement in storage (Duff & Logie, 2001), it is not the substantial drop predicted by the common resource model. Overall, contemporary research has pretty well established that executive WM and STM capacities are not necessarily limited to a common pool of general resources.

Nonetheless, there appear to be some shared general resources. Neuropsychological evidence supports this conclusion: STM storage and executive WM operate from separate structures in the brain, but are directly connected and highly interactive when information in STM storage is being processed (see Chapter 3). Storage seems to suffer most when it is neglected because the individual does not intermittently rehearse the material being held there.

Furthermore, other cognitive factors clearly impact capacity (see Chapter 2). For example, the ability to control attention and inhibit interference and the extent of long-term memory activation all play a role. Processing speed is especially important because slow processing extends the processing interval during which no attention is paid to the items in storage. Processing efficiency and the application of strategies impact retention and performance.

CAUTION

Although the average adult can retain a verbal sequential span of seven chunks in short-term storage, the maximum number of information chunks that can be simultaneously processed by working memory is typically four.

This understanding of WM limitations and capacity has assessment implications. First, span should be measured under minimal or no processing demands and also when there are high processing demands. The first condition is referred to as *simple span* and is measured by such tasks as *digits forward*. Performance on simple span tasks represents STM storage. The latter condition is known as *complex span* and is measured by such tasks as dig-

its backward. Performance on complex spans represents overall WM capacity (Gibson et al., 2012). Another way of framing this is that the STM components should be assessed separately from the WM components. Second, the examinee should be observed and queried in regards to switching and other strategies. Finally, related cognitive processing components, such as processing speed, should be considered.

Regarding overall WM capacity, Cowan (2001) presents extensive convergent evidence that normal adult WM capacity is four chunks. This is distinct from normal adult STM span, which is typically seven words. Cowan views this WM capacity limit as nearly universal, applying across individuals, across modalities, and across levels of expertise. In his view, what varies is the size of the chunks, not the number of chunks. The rule of four chunks applies to normal situations in which individuals are passively attending to information and in which most of the WM processing is automatic. When individuals use a rehearsal strategy to supplement the limited storage function, capacity can be extended to six or seven chunks. Despite Cowan's convincing evidence, some recent studies (reviewed by Verhaeghen, Cerella, & Basak, 2004) have indicated that the typical focus of attention is actually only one item, not four. For example, Oberauer (2002) contends that, at any one time, the focus of attention holds only the single item that is the object of the current or next cognitive operation.

RETENTION INTERVALS

In addition to limits on the amount of information that can be retained and processed, STM and WM are constrained by elapsed time. It has been frequently suggested that the typical retention interval for unrehearsed information is about seven seconds. However, verbal memory traces may fade or decay in as little as two seconds (Baddeley, 1986). It is impossible to specify a normal retention interval in seconds because of confounds that are introduced by rehearsal strategies, long-term retrieval, and processing load. Nonetheless, it is safe to assume that most information that enters STM and is processed in WM is highly degraded within 7 to 15 seconds and completely erased from the short-term store within 20 to 30 seconds (Cowan, 2005; Richardson, 1996). The only exception is when the stored information is being continually processed or rehearsed, in which case the retention interval can be extended indefinitely.

What appears to be a span limitation may actually be a temporal limit. As discussed earlier, individuals can recall about as much as they can articulate, or repeat, in about two seconds (Baddeley, 1986). This finding explains why individuals can recall more short words than long words. The finding also implies that phonological STM holds information for only two seconds unless it is maintained through covert or overt articulatory rehearsal.

While subvocal rehearsal extends the retention interval, it does not necessarily extend the number of items that can be recalled. This has led some researchers (Nairne, 2002) to conclude that span capacity rather than elapsed time is the main factor in forgetting. Nonetheless, there is clearly an inverse relationship between length of interval and length of span. Longer intervals without rehearsal undoubtedly constrict the span (Bayliss, Jarrold, Baddeley, & Gunn, 2003). It appears that extending the duration of the retention interval probably accounts for most of the reduced span performance in complex WM tasks (Conlin, Gathercole, & Adams, 2005).

Distinguishing between STM capacity and WM capacity may clarify some of the discrepancies found across studies. Without rehearsal, passive retention in phonological STM may be as little as two seconds. With rehearsal, this interval can be extended, and, with chunking, the number of items can be increased. That is why it is important to assess different types of WM capacity and to use measures that prevent rehearsal strategies or take strategies into account.

COGNITIVE LOAD THEORY

Cognitive load theory (Van Merriënboer, Kirschner, & Kester, 2003) emphasizes the limited cognitive capacity of WM and how easily WM can become overloaded

during academic learning (T. de Jong, 2010). *Cognitive load* is the processing dimension of WM. Specifically, cognitive load is the proportion of time during which a given processing task occupies WM's focus of attention (Barrouillet, Portrat, & Camos, 2011). Cognitive load does not include the storage dimension of WM but has a direct influence on how much is maintained in storage. As cognitive load increases, there is a corresponding decrease in how much information is retained (Barrouillet et al., 2009). The more that a processing task demands attention, the fewer WM resources are available for rehearsing the information in STM storage components. The processing-storage relationship is bidirectional. Focusing too much on maintaining information through rehearsal can impede processing, slowing it down or causing processing errors.

DON'T FORGET

As cognitive load increases, the amount of information retained in short-term storage decreases.

Cognitive load theory is similar to the general capacity hypothesis discussed earlier. When attention is required for processing, it is not available for the maintenance of memory items and consequently the items fade away. Effective time sharing of attention involves rapid, back-and-forth

switching of attention from processing to maintenance (rehearsal). The crucial variable is the amount of time that is occupied by the processing task (Liefoghe, Barrouillet, Vandierendonck, & Camos, 2008). Memory items are lost when the processing demands are such that the switching cannot occur at all or cannot occur in time to prevent loss of information. Switching in and of itself also adds to cognitive load.

Cognitive load theorists attempt to address cognitive overload in the classroom by identifying causes of high load and by promoting instructional design that minimizes load (T. de Jong, 2010). During instruction and learning, part of the cognitive load is inherent to the content and material to be learned, part is caused by the instructional behaviors of the teacher, and part is created by the learner's internal processing of the information (Kirschner, 2002). The concern is that learning is reduced when too much cognitive load causes loss of information before it can be encoded into LTM. See Chapter 4 for details on what causes high cognitive load in a learning environment.

THE INTEGRATED MODEL OF WORKING MEMORY

The purpose of psychological theories, models, and research is to further understand the functioning of the human brain, in this case WM. For this scientific

information to be useful for real-world assessment and intervention purposes it needs to be operationalized so that WM ability can be effectively and validly measured. The *integrated model* proposed in this text draws from several WM models to frame and organize WM components and functions in a manner that facilitates assessment and intervention. The integrated model, first introduced in Dehn (2008), does not propose any new constructs or structures, but rather integrates models and research so that a whole and comprehensive portrayal of WM is created. Recent research from neuropsychology and neuroscience is considered so that the integrated model is true to what is known about WM neurological structures and functions. The model also incorporates related cognitive functions that need consideration during an assessment of WM (see Chapter 2). Finally, the integrated model is designed to enhance identification of WM weaknesses, deficits, and impairments. The assessment and intervention chapters in this text will reflect the structure of the integrated model.

Short-Term Memory

All of the contemporary models divide WM into processing and storage dimensions. The brief storage components have traditionally been referred to as *short-term memory*. Before the construct of WM was introduced by Baddeley and Hitch (1974), human memory was divided into STM and LTM. Baddeley's model and most other models embed STM within WM, and define it as the storage aspect of WM. This incorporation of STM clouds the independent aspects of STM. The models seldom acknowledge STM as a separate neurological function with its own processing and storage areas. However, to omit STM from separate consideration and assessment is a mistake because it does function automatically at a subconscious level, without necessarily requiring supervision or involvement from WM. It is only when WM is utilizing specific information held in STM that the two systems work in an integrated fashion.

In the integrated model, WM is considered one of many executive functions. As humans speak, the executive function of self-monitoring detects speech and language errors, leading to immediate self-corrections. Yet, language is not considered a subsystem of self-monitoring. Similarly, STM should not be considered only a subsystem of WM. Support for this argument can be found in research that has found weak correlations between short-term auditory memory span and the capacity of the WM central executive (Pennington, Bennetto, McAleer, & Roberts, 1996; Swanson, 1994). Furthermore, factor analysis of memory performance has revealed that STM and WM operate fairly independently of one another (Engle, Tuholski, Laughlin, & Conway, 1999; Swanson & Howell, 2001). Factor

analysis of memory scales has also supported the distinction between STM and WM (Passolunghi & Siegel, 2004).

In the integrated model of WM, STM consists of instantaneous and automated subconscious processes. After perceptual processes have interpreted sensory input, information briefly passes through STM on its way to automatically activating relevant information stored in LTM, such as the meaning of a word that has just been heard. STM also temporarily holds information while it is encoded into LTM. Neither of these automated processes requires conscious participation from WM. Therefore, much of the information handled by STM bypasses WM. Nonetheless, there is nearly constant interaction between executive WM processes and STM storage components. When this type of interaction occurs, it is labeled *verbal WM* or *visual-spatial WM* to distinguish it from the automated, more independent functioning of STM (see Rapid Reference 1.3). In the integrated model, the two STM components are called *phonological STM* and *visual-spatial STM*.

Phonological Short-Term Memory. Phonological STM briefly stores speech-based information in phonological form. Phonological STM continually receives information from auditory sensory stores and automatically activates related items held in long-term storage, such as phonologically similar items. In essence, phonological STM is identical to Baddeley's phonological loop except that subvocal rehearsal is not always a function of phonological STM. Conscious, directed rehearsal efforts are active processes that fall under the purview of executive WM. However, processing occurring below the level of awareness, such as automated rehearsal, is part of phonological STM.

Visual-Spatial Short-Term Memory. This is another STM subcomponent that briefly stores visual (object and color) and spatial (location and direction) information. This subcomponent is the same as Baddeley's visuospatial sketchpad, except that the generation and manipulation of mental images are assigned to WM level. Visual-spatial information is refreshed automatically and continually as objects in the environment change and as the focus of attention changes.

Verbal Working Memory

Verbal WM occurs whenever WM processes utilize verbal information that was just retrieved from LTM or was just perceived and is being temporarily held in phonological STM (see Rapid Reference 1.3). The integrated model does not view recently activated information from LTM as being stored in STM, but rather as

≡ Rapid Reference 1.3 Integrated Model of Working Memory

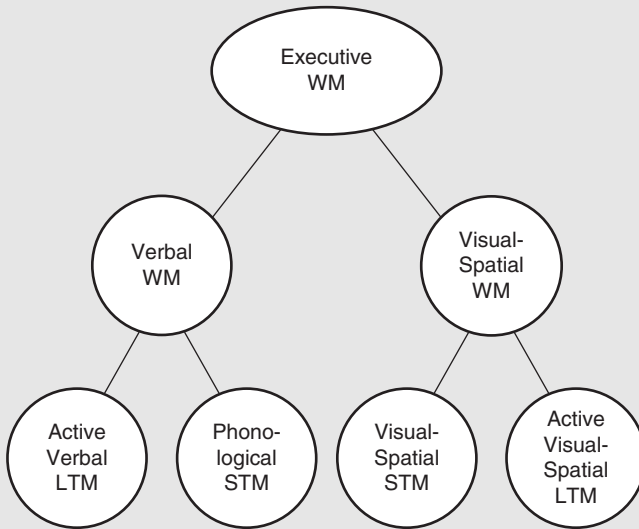


Figure 1.3 Integrated Model of Working Memory

being held in an active region of LTM. This allows WM to quickly go from one recently activated piece of information to another as the focus of attention shifts. This effectively increases WM capacity so that it can process more information concurrently than could be held in very limited STM. For example, Anderson (1983) found that the recently activated LTM region may hold more than 20 LTM units at a time.

Interaction only between WM and STM will not last long, unless the individual has no relevant LTM representations from which to draw, such as someone who has no prior knowledge of a topic. Thus, WM is usually working with both STM and activated LTM units at the same time. An example of this complex interaction would be the processes that occur during reading comprehension. The reader is briefly holding recently decoded words and phrases while at the same time utilizing prior knowledge that has just been activated.

DON'T FORGET

The verbal and visual-spatial processing components of working memory draw information from both short-term storage and recently activated long-term memory representations.

Verbal WM consists of complex WM operations in which analysis, manipulation, and transformation of phonological, auditory, or verbal material take place. One of the primary functions of verbal WM is to extract a meaningful representation that corresponds to the phonological information taken in by phonological STM (Crain, Shankweiler, Macaruso,

& Bar-Shalom, 1990). In contrast to phonological STM, verbal WM is viewed as higher-level, meaning-based processing, whereas phonological STM is simple, passive processing that is more phonologically based.

Visual-Spatial Working Memory

The main distinction between visual-spatial STM and visual-spatial WM is that the STM component involves only passive retention of information, whereas visual-spatial WM adds a processing component, such as reversing the sequence of objects held in STM storage or manipulating an image that was recently activated in LTM. Similar to verbal WM, visual-spatial WM combines information held in STM and LTM. Visual-spatial WM is also involved in creating and manipulating visual images. Until recently, visual-spatial WM was seldom acknowledged as a separate WM component. For a review of empirical support for visual-spatial WM, see Alloway, Gathercole, and Pickering (2006).

Executive Working Memory

According to Baddeley (2003b), there is no verbal or visual-spatial division of WM, except at the STM level. Rather, he views his central executive as modality-free. The functions Baddeley attributes to the central executive are the same as those conducted by verbal and visual-spatial WM in the integrated model. However, the integrated model not only divides WM into verbal and visual-spatial but also retains the concept of a higher-level executive WM that specializes in executive functions not necessarily carried out during specific verbal and visual-spatial processing (see Rapid Reference 1.3).

In the integrated model, executive WM regulates and coordinates all of the cognitive processes that interact with information from either STM or WM. One observable indication that higher-level executive WM is in play is when the

task involves integration of verbal and visual-spatial information. Thus, executive WM is not modality-free but multimodal. Multitasking regardless of the modalities involved would also require executive WM. The unique features of executive WM relative to the modality-specific WM components include the application of strategies that extend the capacity and duration of both the STM and WM subcomponents.

Executive WM operations are both conscious and unconscious. WM theories and research have focused mainly on reportable, conscious functioning. However, a myriad of unconscious specialized operations, such as inhibition, carry out detailed WM functions (Baars & Franklin, 2003). These functions are able to operate subconsciously because they have become automated. Unconscious, automated processing is crucial to successful WM performance because it is believed that automated processing does not draw on the measurable capacity of WM. Nonetheless, automated processes operating below the level of awareness tend to be readily accessible, being called into consciousness whenever effortful processing is required. Operations that were once conscious but became unconscious as their function became automated are the most accessible (Baars & Franklin, 2003).

The other primary executive functions are inhibition, switching, and updating. *Inhibition* is the ability to attend to one stimulus while screening out and suppressing the disruptive effects of automatically generated or retrieved information that is not pertinent to the task at hand. Inhibition also discards previously activated but no longer relevant information and suppresses incorrect responses. In effect, inhibition controls and reduces interference. *Switching*, or shifting, refers to the ability to alternate between different tasks, sets, and operations, such as switching retrieval plans or switching between processing and rehearsing. *Updating* is the ability to keep up with the ongoing flow of information. Updating is a constant process of revision whereby newer, more relevant information replaces old, no longer relevant information (Swanson, Howard, & Saez, 2006).

Interaction With Long-Term Memory

LTM is a vast storehouse of information, much of which is encoded, consolidated, and retrieved through subconscious automated processes. For the most part, LTM can operate independently of WM. For example, when decoding text, a fluent reader automatically retrieves known words and comprehends them without the involvement of WM.

Perhaps the closest interaction between WM and LTM occurs during learning and encoding of information into LTM. For example, when a student listening to a presentation hears new information about a familiar subject, relevant

information in LTM is activated. WM then holds and processes the information, making associations between the new information and activated prior knowledge. The new associations then become encoded into LTM. Another interactive function of WM is to support conscious, effortful retrieval from LTM when the desired information is not automatically and immediately retrieved (Rosen & Engle, 1997). Effortful retrieval from LTM also occurs whenever specific information is demanded of the individual.

The relationship is reciprocal; LTM supports WM as much as WM supports LTM. This is because WM capacity and functioning are affected by the knowledge and skill base in LTM. As knowledge and skills become consolidated and automated in LTM, less processing is required by WM. Moreover, LTM representations may directly enhance short-term and WM spans. When information enters STM, relevant items are immediately activated in LTM. When the information is lost from STM, LTM immediately, automatically, and subconsciously sends cues to WM so that the decayed information can be reconstructed (Nairne, 2002). This interaction explains why individuals are able to remember some newly presented information for longer than a few seconds.

Capacity of Working Memory Operations

In the integrated model of WM, functional capacity, while still limited, may be greater than indicated by span measures. Humans often accomplish WM feats that go beyond predictions based on typical memory spans of only a few items. Incorporating an activated pool of LTM items greatly expands the amount of information available to WM. However, the WM processing dimension still has very limited capacity. Consistent with Cowan's embedded process model, the integrated model proposes that simultaneous processing in WM is limited to approximately four units of information in a typical adult and perhaps only one or two units in a young child. However, the size of the units or chunks may vary, depending on the content and the individual's level of expertise.

The integrated model also proposes that WM performance is determined by how effectively the individual utilizes his innate capacity. For example, the development of automaticity or expertise in a particular skill or content area will enhance WM performance by increasing the size of the information chunks that are manipulated. Applying effective memory strategies from simple rehearsal to more elaborate strategies will also enhance WM performance.

Finally, the integrated model recognizes the influence of cognitive load on WM processing and short-term retention. As cognitive load increases, the amount of information that can be concurrently retained diminishes. Only frequent switching between processing and maintaining items in storage can ameliorate this effect.

SUMMARY

In the integrated model of WM proposed herein, STM, WM, and LTM are all distinct but interrelated memory systems. WM is the interface between STM and LTM, working both with units temporarily retained in STM and with recently activated units from LTM (Rose & Craik, 2012; Unsworth & Engle, 2007). At any point in time, the focus of WM might be material from short-term storage, elements from long-term storage, or a combination of the two.

Without assistance or management from WM, STM and LTM can both function independently. STM can retain information without involvement from WM, although the retention interval is quite short. STM also automatically and independently activates relevant information in LTM. For its part, LTM can automatically and independently encode information briefly held in STM, and it can also function independently in activating and retrieving information.

However, whenever temporarily held or recently activated information requires effortful, conscious processing, WM clearly comes into play. When the interaction is primarily verbal, it is considered verbal WM; when the interaction is primarily visual-spatial, it is considered visual-spatial WM; and when higher-level multi-modal executive functions come into play, it is considered executive WM.

TEST YOURSELF

- 1. Which component from Baddeley's theory is the most difficult to measure?**
 - (a) Phonological loop
 - (b) Visuospatial sketchpad
 - (c) Central executive
 - (d) Episodic buffer
- 2. What is thought to expand the capacity of working memory the most?**
 - (a) Activated pool of LTM items
 - (b) Rehearsal
 - (c) Switching
 - (d) Cognitive load
- 3. In a normal adult how many chunks of information can be managed simultaneously?**
 - (a) 2
 - (b) 4
 - (c) 7
 - (d) It depends on rehearsal.

- 4. Consistent with the integrated model, verbal working memory should be measured separately from phonological short-term memory.**
- (a) True
 - (b) False
- 5. Which is not a key function of executive working memory?**
- (a) Updating
 - (b) Shifting
 - (c) Monitoring
 - (d) Inhibition
- 6. Regardless of age, the length of an individual's phonological STM span is limited to how many monosyllable words the individual can articulate within ___ seconds.**
- (a) 2
 - (b) 5
 - (c) 7
 - (d) 15

Answers: 1. d; 2. a; 3. b; 4. True; 5. c; 6. a