12

Measuring Individual Differences in Working Memory Capacity and Attention Control and Their Contribution to Language Comprehension

Alexander P. Burgoyne, Jessie D. Martin, Cody A. Mashburn, Jason S. Tsukahara, Christopher Draheim, and Randall W. Engle

12.1 Introduction

Imagine you are sitting at a coffee shop talking with a friend. The environment is replete with distractions, from the barista calling out orders to the espresso machine noisily letting off steam. To understand your friend, you must focus on what they are trying to say while (a) preventing your attention from being captured by these distractions and (b) maintaining the gist of what your friend has said in the midst of this sensory and cognitive maelstrom. Of course, this situation is not unique to the coffee shop. Everyday life is filled with distractions and interference, both from the external environment (e.g., receiving a text message) and from internal sources (e.g., thinking about lunch). A ubiquitous challenge, then, is keeping a running gist of the task you are performing while ignoring or suppressing task-relevant and irrelevant distractors. As it turns out, individual differences in these cognitive abilities play an important role in explaining individual differences in language comprehension.

In this chapter, we discuss the measurement of working memory capacity and attention control. First, we examine the origins of complex span measures of working memory capacity, which were created to better

This work was supported by grants from the Office of Naval Research (N00173–18-S-BA01; N00014–12-1-1011) to Randall W. Engle.

understand the cognitive processes underpinning language comprehension. We then discuss the executive attention theory of working memory, which places attention control at the center of individual differences in working memory capacity and fluid intelligence. Next, we describe the relationship between working memory capacity, attention control, and language comprehension, and discuss how maintenance and disengagement – two functions supported by the control of attention – contribute to performance across a range of cognitive tasks. Afterward, we discuss challenges associated with measuring working memory capacity and attention control and identify factors that threaten the construct and criterion validity of these measures. We also detail the steps our laboratory has taken to refine the measurement of these cognitive constructs. We close by providing practical recommendations and resources to researchers who wish to use measures of working memory capacity and attention control in their work.

12.2 The Origins of Complex Span Measures of Working Memory Capacity

Researchers recognized early on that language comprehension requires the short-term storage of information (Kintsch & van Dijk, 1978; Perfetti & Lesgold, 1977). For example, in order to understand the referent of the pronoun “he” in the sentence “Although the doctor was playing golf, he still checked his phone often,” one must recall “the doctor” from the previous clause. The successful integration of information across words and sentences requires that information is not immediately forgotten, but rather is retained at the surface level for a short while and at the gist level for a considerably longer period. These memories must be brought to bear on subsequent comprehension processes to allow what Gernsbacher (1990) called “structure building” at the utterance and sentence level in the short term and at the paragraph and higher level in the longer term. Thus, the interdependence of semantic, syntactic, and contextual information during language processing led theorists to posit that short-term memory played an important role in comprehension. It followed that if short-term memory was indeed critical, individual differences in short-term memory should predict individual differences in language comprehension.

And yet, time and time again, researchers found that short-term memory was largely unrelated to reading or listening comprehension (Hunt et al., 1973; Jackson & McClelland, 1979; Perfetti & Goldman, 1976). As recounted by Daneman and Merikle (1996), “simple span” measures of short-term memory rarely predicted reading comprehension, except when the sample consisted of young children or severely impaired readers (Farnham-Diggory & Gregg, 1975; Rizzo, 1939). This finding was so problematic that it led Crowder (1982) to petition psychologists to abandon the notion of short-term memory. It led Daneman and Carpenter (1980) to argue that language comprehension required more than passive short-term storage; it also required the active maintenance and manipulation of information – in other words, an interplay between the *storage* of prior information and the *processing* of new information. This dual requirement of storage and processing was not adequately captured by simple span short-term memory tasks, they argued, because such tasks merely presented a series of items (e.g., digits) for participants to recall. To address this limitation, Daneman and Carpenter (1980) created the first *complex span* measures of working memory capacity, the reading span and listening span tests.

In the reading span test (Daneman & Carpenter, 1980), participants read aloud a series of sentences, verified whether they made sense, and then recalled the final word of each sentence. The measure of performance was the number of sentences the participant could read while recalling the final words in the correct order. The listening span test was an auditory facsimile of the reading span test; participants listened to sentences instead of reading them aloud. Both complex span tasks interleaved the presentation of memoranda with a secondary task, requiring a trade-off between information storage (e.g., remembering the final words) and processing (e.g., interpreting the sentences). What's more, both components of the tasks used verbal stimuli, which increased the likelihood that the processing and storage subtasks interfered with one another (Hale et al., 1996) and tapped the domain-specific demands of language comprehension.

Daneman and Carpenter (1980) found that performance on the reading span and listening span tests predicted verbal SAT scores and two other reading comprehension measures, fact retrieval and pronominal reference. By contrast, they found that word span and digit span measures of short-term memory did not predict reading comprehension. This pattern of results provided early evidence that the "active ingredient" in measures of working memory capacity was not simple short-term storage, but instead, the ability to control attention to successfully coordinate storage and processing subtasks.

Sixteen years later, a meta-analysis of 77 studies (Daneman & Merikle, 1996) affirmed that complex span measures predict reading comprehension better than short-term storage measures do, and furthermore, that complex span tasks with *verbal* stimuli evoke the strongest relationships between working memory capacity and reading comprehension, likely due to the linguistic processing demands shared across predictor and criterion tasks. Nevertheless, even complex span tasks that do not use verbal stimuli for the processing subtask predict individual differences in reading comprehension (Turner & Engle, 1989), suggesting that these tasks tap a domain-general ability that is important for language processing.

Daneman and Carpenter's (1980) complex span tasks had an enormous impact on the field. Their article, which has been cited over 8,000 times according to Google Scholar, served as a guide for psychometricians to develop other complex span tasks, including *operation span* (Unsworth et al., 2005), *symmetry span* (Unsworth et al., 2009), and *rotation span* (Shah & Miyake, 1996), which we discuss in detail later on. In turn, complex span measures of working memory capacity have been shown to predict individual differences in reading, writing, and speaking ability, as well as following directions (Bock & Miller, 1991; Daneman & Green, 1986; Engle et al., 1991; Gathercole & Baddeley, 1993). Finally, Daneman and Carpenter's (1980) work gave rise to the *capacity theory of comprehension* (Just & Carpenter, 1992), which holds that working memory is essential for language comprehension because it facilitates the resolving of syntactic ambiguity via the maintenance of multiple interpretations, and supports syntactic modularity via the interaction of syntactic and pragmatic information.

For this chapter on the measurement of individual differences, it is worth discussing how complex span tasks fit within the broader theoretical framework of working memory capacity and attention control. Below, we provide a description of what we mean by these terms, followed by a discussion of further evidence for the executive attention view, which places attention control at the center of individual differences in working memory capacity and fluid intelligence.

12.3 The Executive Attention View of Working Memory Capacity

Working memory refers to the cognitive system responsible for the temporary maintenance and manipulation of information in a highly accessible state (Baddeley, 1992). The working memory system comprises a controlled attention component and a short-term storage component, or components. Our view of the working memory system – the executive attention view (Kane & Engle, 2002) – emphasizes the role of attention control, which we define as the domain-general ability to maintain focus on task-relevant information while preventing attentional capture by task-irrelevant thoughts and events.

In Baddeley and Hitch's (1974) classic model of the working memory system (refer back to Figure 2.1), the controlled attention component is called the *central executive*. It is responsible for coordinating the flow of information between short-term storage components in a goal-driven manner (Baddeley, 1992). The *visuospatial sketch pad*, on the other hand, is responsible for the storage of visual information, such as mental imagery. Finally, the *phonological loop* is responsible for the storage of verbal and auditory information, such as speech. What is important for the present purposes is not the exact specifications of Baddeley and Hitch's (1974) model, or subsequent models (Baddeley, 2002), but rather the idea that working memory involves the interplay between (1) controlled attention and (2) short-term memory.

Whereas *working memory* refers to the cognitive system responsible for the temporary maintenance and manipulation of information, *working memory capacity* refers to the measurement of an individual's ability to use the working memory system. Working memory capacity is often measured using *complex span tasks*: dual tasks that require participants to simultaneously store and process information. Although working memory capacity is often indexed in terms of "number of items recalled," short-term storage only tells part of the story. This is because people who are better able to flexibly allocate attention to the storage and processing subtasks perform better on working memory tasks than those who are more susceptible to distraction and interference. In fact, evidence suggests that the controlled attention component of working memory plays a large role in explaining its relationships with a range of outcomes and abilities (Kane & Engle, 2002), including language comprehension.

Early evidence for the executive attention view was provided by studies that found differences between high and low working memory capacity participants on tasks that demanded controlled attention but placed little burden on short-term memory. For example, in the *antisaccade task*, participants must resist the urge to look at a flashing cue on one side of the screen, and instead rapidly look toward the opposite side of the screen. The task is challenging because it requires inhibiting a reflexive response: looking at a highly salient visual stimulus. Adding to the difficulty, participants cannot simply ignore the flashing cue, because it indicates the side of the screen they should look away from. In a sample of 203 participants, Kane et al. (2001) found that people with higher working memory capacity made fewer errors on the antisaccade task (i.e., looking at the cue instead of away from it), and were quicker to recover when they looked in the wrong direction. Because the antisaccade task does not burden participants with lists of items to remember, but does require controlled attention, Kane et al.'s (2001) results suggest that individuals' working memory capacity is closely linked to the functioning of the central executive.

As another example, in a dichotic listening task, participants are presented two auditory streams, one to each ear, and must repeat aloud the messages presented to one ear while ignoring the messages presented to the other. At some point, the participant's name is surreptitiously presented to the unattended ear. Later on, the participant is asked whether they heard anything unusual. Around one-third of people report hearing their name in the unattended ear, an effect which has been termed the "cocktail party" phenomenon (Moray, 1959) Critically, low working memory capacity participants were significantly more likely to report hearing their name than high working memory capacity participants (Conway et al., 2001). This suggests that people with high working memory capacity are better able

Comp. by: P. Udayasankar Stage: Proof Chapter No.: 12 Title Name: SchwieterandWen
Date:18/2/22 Time:08:55:36 Page Number: 252

to control their attention to ignore task-irrelevant distractors, and may explain why some people perform better than others in the coffee shop scenario described in the beginning of this chapter.

Additional evidence for the executive attention view is provided by studies that use latent variables to model relationships between cognitive constructs. Latent variables are unobserved variables that capture variance common to a set of indicators (e.g., performance measures on different tasks). Latent variable analyses such as confirmatory factor analysis and structural equation modeling provide a number of advantages relative to other statistical approaches. For instance, latent variable analyses allow researchers to draw conclusions about cognitive constructs – the hypothesized source of the shared variance among a set of measures – as opposed to drawing conclusions about specific measures, which may only capture a slice of the cognitive construct of interest. Furthermore, latent variables are theoretically free of measurement error, which attenuates relationships (Kline, 2015). As we discuss in the "Practical Recommendations" section below, latent variable analyses require large samples (e.g., 250 or more participants) and multiple measures per construct (e.g., 3 or more measures). When the proper conditions are met, however, latent variable analyses can be a powerful tool for elucidating relationships between constructs.

Using latent variable analyses, Engle et al. (1999) found that controlled attention drives working memory capacity's relationship with *fluid intelligence* (i.e., reasoning ability). First, Engle et al. (1999) established that working memory capacity and short-term storage were dissociable at the latent level. The two constructs shared approximately 46 percent of their reliable variance – a substantial amount, but considerably less than 100 percent. More importantly, they found that working memory capacity contributed independently to fluid intelligence after accounting for short-term storage. While the predictive path from short-term storage to fluid intelligence was near zero and nonsignificant, the path from working memory capacity to fluid intelligence was substantial and significant (Figure 12.1). This indicates that the controlled attention component of working memory contributes to fluid intelligence above and beyond short-term memory.

Twenty-one years later, Draheim et al. (2021) corroborated and extended this finding by analyzing relationships between working memory capacity, attention control, and fluid intelligence. Whereas Engle et al. (1999) estimated the contribution of controlled attention by partialling out variance in working memory capacity attributable to short-term storage, Draheim et al. (2021) measured attention control directly using a battery of new-and-improved attention tasks, described in greater detail in the "Measuring Attention Control" section below. Using structural equation modeling, Draheim et al. (2021) found that attention control mediated the relationship between working memory capacity and fluid intelligence. That is, the once-significant relationship between working memory

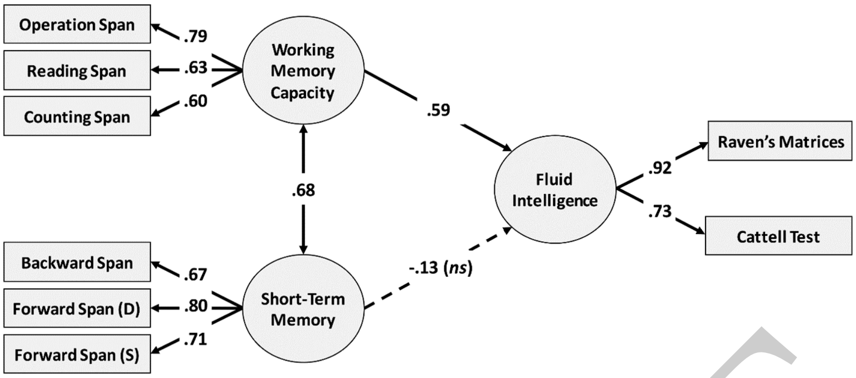

Figure 12.1 The substantial and significant contribution of working memory capacity to fluid intelligence after accounting for short-term memory
Note: latent factors are depicted as ovals and observed measures are depicted as rectangles. Working memory capacity and short-term memory were strongly correlated (.68). The contribution of working memory capacity to fluid intelligence (.59) was substantial and significant after accounting for short-term memory. By contrast, the contribution of short-term memory to fluid intelligence (-.13, ns) was non-significant after accounting for working memory capacity. This pattern of results suggests that the controlled attention component of working memory contributes to fluid intelligence above and beyond short-term memory. Adapted from Engle et al. (1999).

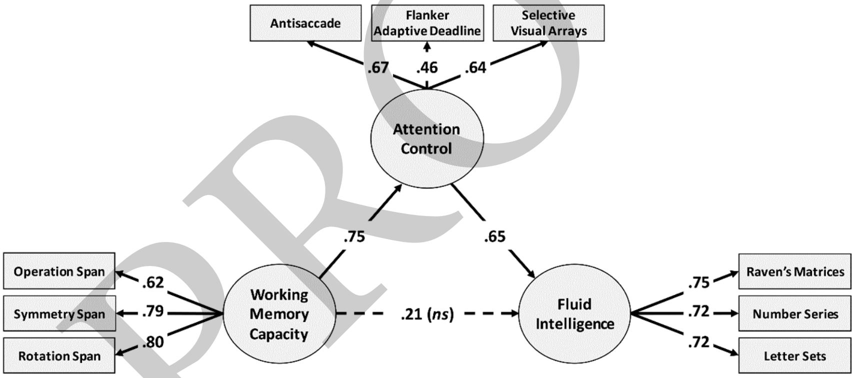

Figure 12.2 A structural equation model depicting attention control fully mediating the relationship between working memory capacity and fluid intelligence.
Working memory capacity was strongly related to attention control (.75), and attention control was strongly related to fluid intelligence (.65). Working memory capacity did not predict fluid intelligence (.21, *ns*) after accounting for attention control. The results suggest that attention control drives the relationship between working memory capacity and fluid intelligence. Adapted from Draheim et al. (2021).

capacity and fluid intelligence was no longer significant after accounting for attention control (Figure 12.2). This finding extends the work of Engle et al. (1999) by showing that individual differences in attention control can fully explain the relationship between working memory capacity and fluid intelligence.

12.4 Working Memory Capacity, Attention Control, and Language Comprehension

Returning to the domain of language, we see that evidence also suggests that the controlled attention component of working memory contributes to individual differences in comprehension. For example, Swanson and Ashbaker (2000) found that working memory capacity significantly predicted reading comprehension and word recognition performance in children with learning disabilities, even after accounting for short-term memory. Across a series of hierarchical regression analyses, the incremental validity of working memory capacity above and beyond short-term memory ranged from 5 percent to 27 percent. For comparison, the incremental validity of short-term memory above and beyond working memory capacity ranged from 1 percent to 7 percent. Because the unique variance in working memory capacity after accounting for short-term storage represents controlled attention, this pattern of results indicates that attention control drives the relationship between working memory capacity and language comprehension (see also Engel de Abreu et al., 2011).

Further evidence for the role of attention control in supporting language comprehension is provided by experiments that burden the central executive with a distractor task and reveal concomitant decreases in language comprehension. For example, Waters et al. (1987) had participants maintain a random sequence of six digits while reading sentences of varying syntactic complexity. They found that burdening the central executive significantly impaired comprehension of syntactically complex sentences. As discussed by Caplan and Waters (1999), concurrent digit load primarily affects the comprehension of syntactically complex sentences when the presentation of one set of stimuli interrupts the presentation of the other. This suggests that attentional shifts induced by a secondary task may interfere with efforts to structure sentences syntactically or interpret their meaning (Caplan & Waters, 1999).

Given that the relationship between working memory capacity and comprehension is partly attributable to controlled attention, some language researchers have attempted to measure attention control directly, rather than indirectly by partialling out variance in working memory capacity attributable to short-term storage or manipulating it by burdening the central executive. However, the measurement of individual differences in attention control poses its own challenges due to psychometric limitations, as we discuss in the section "Measuring Attention Control." Nevertheless, researchers have found that measures of attention control predict individual differences in language abilities.

For example, McVay and Kane (2012) found that latent variables representing attention control and reading comprehension ability were strongly correlated in a sample of over 200 participants. To explore the mechanism by which attention control contributed to comprehension, McVay and Kane

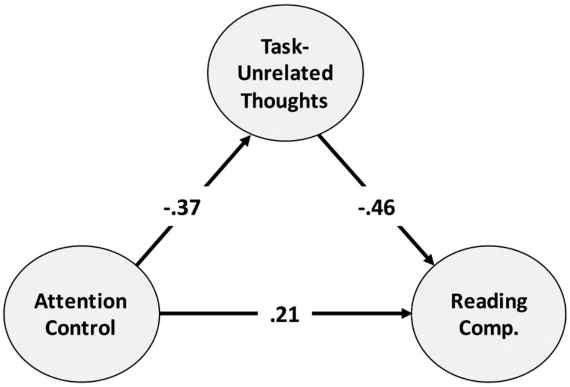

Figure 12.3 A structural equation model depicting task-unrelated thoughts partially mediating the relationship between attention control and reading comprehension
Attention control was negatively related to the number of task-unrelated thoughts (-.37), and task-unrelated thoughts were negatively related to reading comprehension (-.46). The path from attention control to reading comprehension (.21) remained significant after accounting for task-unrelated thoughts, indicating that attention control contributed to reading comprehension even after accounting for mind wandering.
Adapted from McVay and Kane (2012).

(2012) had participants report instances of "task-unrelated thoughts," or mind wandering, during task performance. Attention control was negatively correlated with the frequency of task-unrelated thoughts, indicating that participants with greater attention control were better able to maintain task focus and were less susceptible to distractions. Furthermore, the relationship between attention control and reading comprehension was partially mediated by task-unrelated thoughts (Figure 12.3). In other words, mind wandering partly explained the relationship between attention control and reading comprehension. That said, the direct path from attention control to reading comprehension remained significant even after accounting for task-unrelated thoughts, suggesting that mind wandering only captured part of the covariance between attention control and comprehension. McVay and Kane (2012) speculated that this unexplained covariance between attention control and reading comprehension may represent effective competition resolution, which is measured by tests of attention control and required when readers encounter ambiguity in a passage of text.

As another example, Blankenship et al. (2019) examined the development of infants' attentional abilities and their relationship to reading achievement at age 6 in a longitudinal study of 157 children. Blankenship et al. (2019) measured the attention abilities of 5-month-old infants by showing them a 45 second video clip from Sesame Street. They counted the number of times the infants shifted their gaze, and the longest duration they looked at the video. Blankenship et al. (2019) found that infants' attentional abilities predicted their executive functioning five months later, as measured by the A-not-B task (i.e., an "updating" test in which infants are challenged to find a toy hidden in a new location; to do so, they must avoid perseverating on a previously learned location). They also found that differences in executive functioning were reliable and displayed continuity with age, such that each subsequent measure (obtained at ages 3, 4, and 6 using span tasks and tests of attention control) was significantly related to the previous developmental measure. Critically, the relationship between attentional abilities in infancy and reading achievement at age 6 was mediated by executive functioning across development. This result held even after controlling for verbal intelligence, suggesting that individual differences in domain-general attentional abilities can be detected early on, and contribute to reading achievement above and beyond domain-specific language abilities.

12.5 Maintenance and Disengagement

Why does attention control contribute to performance on working memory, fluid intelligence, and language comprehension tasks? In our theoretical framework, attention control is necessary for performing two distinct but complementary cognitive functions that are important to a wide range of tasks. Those two functions are maintenance and disengagement (Burgoyne & Engle, 2020; Shipstead et al., 2016). *Maintenance* refers to the cognitive operations that support keeping track of information, particularly amid distraction and interference. For example, maintenance is required when building the gist of a complex story told by your friend in the coffee shop among lots of distractions and interruptions. Sources of interference can include task-irrelevant thoughts, as well as external events that threaten to capture attention. *Disengagement*, on the other hand, is responsible for removing no-longer-relevant information from active processing, and flagging it for nonretrieval. For example, one must disengage from irrelevant information that was processed during the interruptions in your friend's story. We think most tasks require both information maintenance and disengagement, but the extent to which each is important depends on the cognitive demands of the task at hand (Figure 12.4).

For example, in complex span working memory tests, maintenance plays a critical role because the performer must keep track of memoranda while completing secondary processing tasks. Disengagement seems less important than maintenance in complex span tasks; however, the performer must still disengage from memoranda from prior trials and the processing subtasks to perform well. By contrast, in fluid intelligence tests such as Raven's matrices (Raven & Court, 1998) or number series (Thurstone, 1938), we think disengagement plays a larger role than maintenance. Many fluid intelligence tasks challenge participants to discover relationships or abstract rules among stimuli. As participants rule out disproven hypotheses, they must prevent these incorrect hypotheses from being reretrieved, reentering the focus of attention, and interfering with the discovery of novel

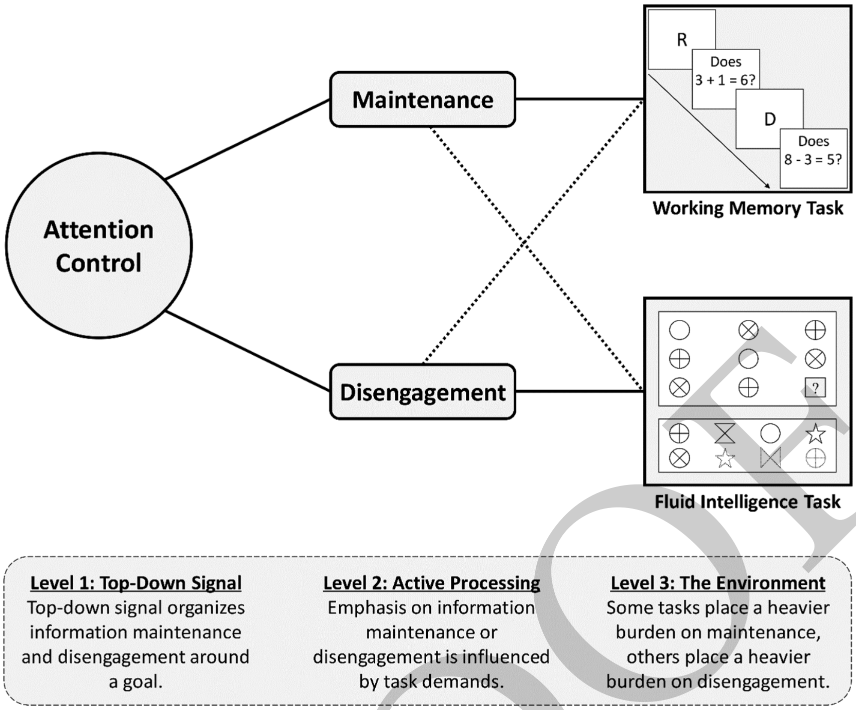

Figure 12.4 Attention control supports information maintenance and disengagement in service of complex cognition
Adapted from Burgoyne and Engle (2020).

solutions (Burgoyne et al., 2019b; Hambrick & Altmann, 2015). Maintenance, on the other hand, may help fluid intelligence test takers keep track of information used to generate novel hypotheses (see Burgoyne et al., 2019b). With respect to language comprehension, maintenance appears to make a substantial contribution because readers must keep track of previous information to contextualize new information, and must maintain multiple interpretations of ambiguous sentences until they are resolved in the service of structure building (Gernsbacher, 1990). Disengagement also appears to play a role; once ambiguity in a sentence has been resolved, such as after a garden-path sentence, the incorrect interpretation of that sentence should be removed from further consideration.

In a recent large-scale study, Martin et al. (2020) estimated the contribution of information maintenance and disengagement to reading comprehension and second-language vocabulary learning. Martin et al. (2020) had 567 young adults (ages 18–35) complete tests of working memory capacity, memory updating, and fluid intelligence. Using structural equation modeling, Martin et al. (2020) partitioned variance in performance on these tasks into latent factors representing maintenance and disengagement. The models revealed that maintenance and disengagement were statistically separable at the latent level. Moreover, each made substantial and significant contributions to reading comprehension and second-language vocabulary learning; together, they accounted for 58 percent of the variance in reading comprehension and 61 percent of the variance in second-language vocabulary learning.

12.6 Measuring Attention Control

Although the preceding results suggest that attention control contributes to language comprehension, psychometric limitations have posed a challenge for researchers attempting to directly measure individual differences in attention control. These limitations were shown clearly at the latent level by Friedman and Miyake (2004), who had 220 undergraduates complete nine tasks designed to measure three attentional functions (inhibiting a prepotent response, resisting distraction, and resisting proactive interference). Friedman and Miyake (2004) found that most of the measures were unreliable, with an average internal consistency below .60. Because unreliability attenuates correlations, it is perhaps unsurprising that the measures correlated weakly with each other (only one was above $r = .18$), and that the average factor loading was below .40. In light of these results, Friedman and Miyake (2004) suggested that researchers develop new tests of attention control with greater reliability, process purity, and sensitivity to individual differences.

Despite this suggestion, many of the attention tasks used by Friedman and Miyake (2004) are still used today. The use of psychometrically unsound tasks has led some to conclude that measures of attention reflect task-specific factors and not an underlying unitary ability (Kramer et al., 1994). Others have argued that it is difficult to draw conclusions about the unity or diversity of attention control as a cognitive construct in the presence of psychometric limitations such as unreliability (Paap & Sawi, 2016), contamination by processing speed, strategy, semantic memory, and speed-accuracy trade-offs (Draheim et al., 2019; Hedge et al., 2020), and a small effect size to noise ratio (Rouder & Haaf, 2019).

We have argued that problems affecting the measurement of attention control are largely due to the use of response time difference scores, which reduce reliability and induce contamination by processing speed and speed-accuracy trade-offs (Draheim et al., 2019; 2021). *Differences scores* use a subtraction methodology; an individual's performance in one condition is subtracted from their performance in another condition. For example, in the Stroop task, participants must indicate the color a word is printed in, not the color it refers to (Stroop, 1935). Trials can be *congruent* (e.g., "RED" in red ink) or *incongruent* (e.g., "BLUE" in red ink). Performance on incongruent trials is hypothesized to require controlled attention, because participants must resolve conflict between the word and the color it is printed in. By contrast, performance on congruent trials requires largely nonattentional processes, given the lack of conflict resolution required and the automaticity of reading

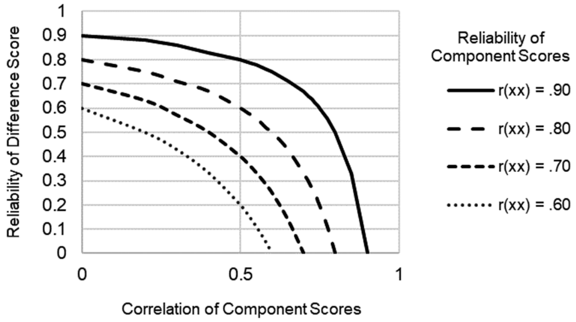

Figure 12.5 Reliability of a difference score (*Y*-axis) decreasing as the correlation between the component scores increases (*X*-axis)
Each line represents the reliability of the difference score when the reliabilities of the component scores are set to *r*(xx) = .60, .70, .80, or .90.

(see MacLeod, 1991, for a review). The difference between performance on incongruent and congruent trials is thought to reflect attention-specific variance, and for this reason many attention tasks use difference scores between performance on conditions requiring controlled attention and conditions thought to reflect largely automatic processes.

The subtraction methodology appears to be a great tool for experimental researchers (see Chiou & Spreng, 1996), but the use of difference scores in individual differences research has been denounced by psychometricians for over half a century (Cronbach & Furby, 1970). Many researchers have noted that difference scores are poorly suited for correlational work because they are often unreliable and minimize between-subjects variance (Ackerman & Hambrick, 2020; Draheim et al., 2016; 2019; Friedman & Miyake, 2004; Hedge et al., 2018). Difference scores are less reliable than their *component scores* (i.e., the performance measures from each condition used to calculate the difference score) in all practical situations because subtraction removes the shared – and therefore reliable – variance of the component scores but preserves the error variance. As shown in Figure 12.5, the unreliability of a difference score depends on the reliability of its components and how strongly those components are correlated. For attention tasks, congruent and incongruent trials are typically highly reliable (e.g., .90) and strongly correlated (e.g., .80), leading to unreliable difference scores and subsequently poor validity (Draheim et al., 2021; Hedge et al., 2018; Paap & Sawi, 2016).

With these issues in mind, we recently developed new-and-improved tasks to measure attention control (Draheim et al., 2021). We administered ten attention tasks to over 400 participants, including "classic" tasks (e.g., the antisaccade task), modified tasks (e.g., the Stroop task with an adaptive response deadline), and new tasks (e.g., the sustained-attention-to-cue task). Our new and modified tasks avoided the use of difference scores. Many of them used an adaptive procedure in which the tasks became easier or more difficult depending on how the participant performed. In these adaptive tasks, we set the converged-upon accuracy rate to be constant across participants, and used the level of task difficulty at which the participant could perform at this accuracy rate as the dependent measure. The new accuracy-based attention tasks were markedly better than classic tasks that relied on difference scores or response times in terms of reliability, intercorrelations, loadings on an attention control factor, and associations with fluid intelligence. As we noted above, using these new tasks, we found that attention control fully mediated the relationship between working memory capacity and fluid intelligence at the latent level, a result that could not be attributed to processing speed. Furthermore, the results suggested that attention control *is* a unitary ability, when measured using psychometrically sound tasks.

12.7 Measuring Working Memory Capacity

The measurement of working memory capacity is considerably less contentious than that of attention control, with several psychometrically sound tasks available to researchers. The strong reliability and criterion validity of these measures is largely attributable to the tasks being designed for individual differences research and scored without using response times or difference scores. That said, there are at least three ongoing issues pertaining to the measurement of working memory capacity to consider, including whether tasks are interchangeable, whether they are appropriate for lower- and higher-ability samples, and whether administration time can be reduced without loss in reliability or criterion validity.

Although researchers use a variety of tasks to draw conclusions about working memory capacity, these conclusions may differ depending on the tasks used to measure it. For example, while complex span tasks are popular, so are the *n*-back and running span tasks. In the *n*back, participants are presented a continual series of stimuli (e.g., letters) and must respond when the current stimulus is identical to the stimulus presented *N* trials ago. In the running span task, participants are presented a series of stimuli and must recall the last *x* number of stimuli in the order they were presented.

Despite ostensibly measuring the same construct, a meta-analysis by Redick and Lindsey (2013) revealed that performance on complex span tasks correlated weakly with performance on *n*-back tasks ($r = .20$) – the two measures shared only 4 percent of their variance. In another study, this time using latent variable analyses, Harrison (2017) found that complex span and *n*-back measures loaded onto separate factors that shared less than one-quarter of their reliable variance. Moreover, both factors accounted for unique variance in fluid intelligence. Harrison (2017) also found that *n*-back tasks with a larger stimulus pool loaded onto a separate factor than *n*-back with a smaller stimulus pool, which in turn affected their relationships with complex span performance and fluid intelligence.

These results suggest that measures of working memory capacity based on complex span and *n*-back tasks may not reflect the same construct or source of variance, and that task-specific factors may play a role in explaining contradictory results across studies that use the *n*-back.

On the other hand, Broadway and Engle (2010) found that running span and complex span performance was strongly correlated, and that both measures had nearly equivalent relationships with fluid intelligence. Furthermore, Broadway and Engle (2010) found that these relationships were largely invariant to task-specific factors in the running span task, such as the presentation rate and whether the participant knew the *set size* (i.e., the number of items to be remembered) in advance. Taken together, these studies indicate that working memory capacity measures are not always interchangeable. A robust assessment of working memory capacity should therefore include more than one type of task, as we discuss in the "Practical Recommendations" section below.

Another consideration is the match between the difficulty of the task and the ability level of the population of interest. To shed light on this issue, Draheim et al. (2018) used item response theory to analyze three complex span tasks: operation span, symmetry span, and rotation span. The analyses revealed that the standard operation span task was poor at differentiating between high- or even average-ability individuals, in part because there were *ceiling effects* (i.e., performance at or near 100 percent) on trials with lower set sizes. As a result, operation span performance and fluid intelligence were not significantly correlated among the top third of performers unless larger set sizes were added to make the task more difficult. For comparison, the standard rotation and symmetry span tasks were much better at distinguishing between average- and high-ability individuals, although they also benefited from adding larger set sizes. Because the smaller set sizes used in these three complex span tasks only tapped variance among the worst performers, Draheim et al. (2018) concluded that in many cases they could be removed to reduce administration time.

A final consideration is that working memory capacity tasks are time-consuming. A battery of three standard complex span tasks takes over an hour to administer. Recent efforts to shorten these tasks by reducing practice time, removing smaller set sizes, and reducing the number of trials have been relatively successful. For example, Foster et al. (2015) and Oswald et al. (2015) used different approaches to shorten complex span tasks but converged on a similar conclusion: although shortening the tasks reduced their internal consistency reliability, it decreased their administration time by 20–40 percent and left their criterion validity largely intact. Nevertheless, three shortened complex span tasks still require over 40 minutes to complete, compared to roughly 30 minutes for attention control and 25 minutes for fluid intelligence tasks (Draheim et al., 2021). As such, complex span tasks could benefit from further efforts to shorten their administration time, perhaps by making them adaptive in difficulty on a trial-by-trial basis.

Comp. by: P. Udayasankar Stage: Proof Chapter No.: 12 Title Name: SchwieterandWen
Date:18/2/22 Time:08:55:37 Page Number: 262

12.8 Practical Recommendations

In this section, we provide practical recommendations to researchers interested in conducting studies of individual differences in cognitive ability. Given that *differential psychology* (i.e., the study of individual differences) is rarely taught in undergraduate- and graduate-level methods courses, and that best practices for differential research are rarely discussed in scientific publications (see Burgoyne et al., 2020), these recommendations may not be obvious to the uninitiated differential researcher.

12.8.1 Carefully Consider Whether the Cognitive Tasks You Administer Will Reflect the Cognitive Construct You Intend to Measure Given Your Population of Interest

It is all too easy to select a task described as a measure of a cognitive construct, administer it to a sample, and assume you are properly measuring that construct or ability. While this may sometimes be the case, researchers should consider whether the demographics of the sample the task was developed and validated for are comparable to the researcher's sample of interest. For example, the same "working memory" task may reflect different abilities when administered to different age groups, such as young adults or children. Other scenarios might not be as obvious, for instance, administering a computerized task to a sample that is not proficient in using a computer, or ensuring that task instructions are fully understood by non-native speakers or those with lower language proficiency.

12.8.2 Ensure Your Sample of Subjects Reflects a Broad Range of Abilities

Measures of cognitive abilities are designed to identify individual differences, so it is critical to include individuals who differ in your sample if you are interested in the population at large. A homogenous sample (e.g., university students enrolled in an introductory psychology course) can result in severe restriction of range, leading to reduced between-subjects variance (i.e., reduced variability across participants) and therefore lower reliability and validity of the measures.

12.8.3 Individual Differences Studies Require Larger Samples Than Typical Experimental Studies

For example, Schönbrodt and Perugini (2013) concluded that correlations for moderate-sized effects do not stabilize until the sample size approaches 250 participants. Determining the actual sample size required for stable correlations is more complex than conducting a power analysis; we suggest

Table 12.1 *Working Memory Tasks Available for Download from https://englelab.gatech.edu*

Task	Versions	Description
Reading Span	Standard Shortened Translated	Participants read sentences and remember a word presented after each sentence. After a series of sentences, participants recall the words in the presented order.
Operation Span	Standard Advanced Shortened Translated	Participants solve simple three-term math problems such as "does $(3 \times 5) - 7 = 8$?", each of which is followed by the presentation of a letter. After a series of items, participants recall the letters in the presented order.
Symmetry Span	Standard Advanced Shortened Translated	Participants judge whether a pattern of black and white squares is symmetrical about the midline. Afterward, they are shown a 4×4 grid with one square depicted in red. After a series of items, participants recall the position of the red squares in the presented order.
Rotation Span	Standard Advanced Shortened Translated	Participants mentally rotate uppercase consonants to the upright position and indicate whether the consonant is mirror-reversed. Afterward, they are presented with a big or small arrow pointing one of eight possible directions. After a series of items, participants recall the arrows in the presented order.
Counting Span	Standard Translated	Participants count the number of target shapes appearing in a display of targets (e.g., blue circles) and non-targets (e.g., green circles). After a series of displays, participants recall the number of targets in each array in the presented order.
Running Span	Standard Translated	Participants are presented with a list of letters and must report the last *x* number of letters in the presented order.

Note: These tasks are available for download at https://englelab.gatech.edu/. Standard = standard administration; Advanced = includes larger set sizes for higher-ability participants; Shortened = reduced administration time; Translated = translated into a language besides English.

consulting Table 12.1 in Schönbrodt and Perugini (2013). Latent variable analyses are problematic when samples are too small to produce a robust and stable correlation matrix.

12.8.4 Avoid Using a Single Task to Measure a Cognitive Ability

Although cognitive tasks are designed to measure processes associated with a particular cognitive ability, any single task will inevitably also reflect a wide variety of extraneous cognitive processes. This creates a problem when researchers attempt to equate scores on a single task to a cognitive ability. Instead, a latent variable approach should be used to capture variance common to a set of tasks. Extraneous processes will not be captured by the latent variable if they are unshared across tasks. As a rule of thumb, at least three tasks per latent construct is advised.

12.8.5 The Particular Set of Tasks Used to Measure a Cognitive Ability Matters

Although a latent variable approach provides a more "process pure" and theoretically meaningful measure of cognitive ability relative to analyses of observed variables, in practice, the reliability and validity of the latent variable will depend heavily on the set of tasks used. For instance, measuring working memory capacity by exclusively using complex span tasks, or *n*-back tasks, will result in somewhat different latent constructs. Even though both sets of tasks are considered working memory capacity tasks, there is evidence that they do not necessarily measure the same thing (Redick & Lindsey, 2013).

12.8.6 For Broad Cognitive Abilities, Use a Heterogeneous Set of Tasks That Reflect Different Domain-Specific Processes

For instance, a set of tasks should tap both verbal and spatial abilities. This ensures that the common variance that is captured by the latent factor is domain-general and does not reflect more domain-specific abilities. Note that one potential limitation of current attention control tasks is that they often rely on visual-spatial processing, and there are not many that require verbal or auditory processing.

12.8.7 Do Not Use Difference Scores to Assess Individual Differences

As we discussed in the "Measuring Attention Control" section, *difference scores* subtract an individual's performance in one condition from their performance in another condition and should be avoided in correlational research. Difference scores are less reliable than their component scores, leading to a poor signal-to-noise ratio and attenuated validity. Furthermore, difference scores are not a necessary requirement to measure individual differences in attention control. In the "Task Downloads" section below, we provide researchers with attention tasks that do not use difference scores.

12.8.8 Do Not Use an Extreme-Groups Design

Extreme-groups design refers to comparing a subset of participants categorized into high- versus low-ability groups. Historically, extreme-groups comparisons have been used to circumvent the need for large samples because they do not require the full continuous range of an ability. However, from our own experience, extreme-groups designs have the potential to confound one ability with another. For instance, because working memory capacity and fluid intelligence are highly correlated, participants that are categorized as high working memory capacity will also have high fluid intelligence. Extreme-groups designs can be used for exploratory purposes or to justify a larger-scale study, but ideally, researchers should use a large sample representing a continuous range of an ability before making strong claims about cognitive mechanisms.

12.8.9 Measure Cognitive Abilities That Are Highly Related to Predictor and Criterion Variables

To draw conclusions about the mechanisms underlying performance in a domain, researchers should include measures that are highly related to the predictor and criterion variables. Simply showing that a cognitive ability is correlated with performance in a domain does not provide strong evidence that the cognitive ability underpins performance in that domain. For instance, if working memory capacity is moderately correlated with a criterion measure, it is likely that fluid intelligence and attention control will also correlate with the criterion measure. Without measuring these constructs, it cannot be determined which construct is most important. What is needed is incremental validity: Does working memory capacity predict the criterion measure above and beyond attention control and fluid intelligence? Granted, it is unfeasible to measure every variable that might be related to the predictor and criterion variables, but researchers should endeavor to rule out theoretically plausible third-variable explanations.

12.8.10 Report Detailed Demographic Information, Task Descriptions, Descriptive Statistics, Reliabilities, and Bivariate Correlations

A detailed description of your sample (e.g., age, sex, level of education, race, ethnicity) helps researchers evaluate the validity and generalizability of your conclusions, and determine whether the tasks you administered were appropriate for the sample under investigation. The same is true for reporting descriptive statistics, reliabilities, and bivariate correlations between measures.

Some cognitive tasks have standardized administration procedures, however, many attention control tasks do not. Laboratories vary widely in the properties of the stimuli that are presented, the proportion of trial types (e.g., congruent vs. incongruent), the total number of trials, and so on. Some of these differences will have a greater impact on the reliability and validity of the task than others. For instance, many attention control tasks may require more trials than are typically administered to attain adequate reliability (Rouder et al., 2019). However, tasks that are adaptive in difficulty may reduce administration time without significant loss in reliability and validity (Draheim et al., 2021). Also, the proportion of trial types interacts with individual differences in cognitive ability in both the Stroop and flanker tasks (Heitz & Engle, 2007; Kane & Engle, 2003). In general, a higher proportion of congruent trials to incongruent trials (e.g., 2:1) is optimal for capturing individual differences in attention control. While our research has shown that the order in which tasks are administered can actually change what the task is measuring (see, e.g., Kane et al., 2001), counterbalancing task order between participants may only add more noise if you do not include the counterbalance variable in the statistical model. Our recommendation is for researchers to think carefully

Table 12.2 *Attention Control Tasks Available for Download from https://englelab.gatech.edu*

Task	Description
Antisaccade	Participants fixate centrally as a flashing cue appears to the side. Participants must rapidly look to the opposite direction to detect a briefly presented letter.
Color Stroop Adaptive Deadline	Participants are shown color words appearing in different colors. On *congruent* trials, the word and its color match ("RED" in red color), whereas on *incongruent* trials, the word and its color do not match ("BLUE" in red color). Participants must report the color the word appears in. This task uses an adaptive response deadline; better performance results in a shorter response deadline, worse performance results in a longer response deadline. The dependent measure is the response deadline duration at the end of the task.
Arrow Flanker Adaptive Deadline	Participants are presented five arrows and must indicate the direction of the central arrow. The flanking arrows can either be *congruent* (←←←←←) or *incongruent* (←←→←←) with the direction of the central arrow. This task uses an adaptive response deadline; better performance results in a shorter response deadline, worse performance results in a longer response deadline. The dependent measure is the response deadline duration at the end of the task.
Selective Visual Arrays	Participants are cued to pay attention to either the red or blue rectangles. Next, they are briefly shown a target array of red and blue rectangles in different orientations. After a visual mask, a test array appears that is either identical to the target array or differs by one item. A white dot appears on one of the items, and the participant must determine whether that item changed or remained the same across the two arrays.
Sustained Attention to Cue	Participants are shown a circle cue that gradually narrows on a target location to be monitored for the presentation of a letter. After a variable wait interval, a centrally-presented asterisk flashes, and a letter surrounded by distractors is briefly presented at the target location. The participant must indicate the letter.

Note: These tasks are available for download at https://englelab.gatech.edu/.

about the order of tasks presented to participants, with consideration of potential spill-over effects across tasks if counterbalancing is not used.

12.9 Task Downloads

Researchers can download working memory capacity and attention control tasks for free from our laboratory website: https://englelab.gatech.edu/. In Tables 12.1 and 12.2, we describe the tasks available for download. Many of the working memory capacity tasks available for download have been translated into languages besides English. We also provide versions of our tasks with shortened administration times, and advanced versions that contain larger set sizes for higher-ability samples. All tasks were programmed in E-Prime.

12.10 Conclusion

In this chapter, we discussed the measurement of working memory capacity and attention control, and how these measures have been used by language researchers to better understand the cognitive processes underpinning comprehension. We also discussed challenges associated with measuring these cognitive abilities, and provided recommendations and resources to researchers interested in conducting studies of individual differences. Although we did not discuss what governs the top-down control of attention, one plausible explanation is that what we attend to is guided in part by the contents of long-term memory and their interaction with environmental cues (see Delaney, 2018, and also Adams & Delaney, this volume). At the very least, such attempts at explaining the top-down control of attention help circumvent an infinite regress of central executives controlling central executives. Research on attention control is evolving rapidly, and we are excited to see where the adoption of more sophisticated statistical techniques and measurement methods take the field.

References

Ackerman, P. L., & Hambrick, D. Z. (2020). A primer on assessing intelligence in laboratory studies. *Intelligence*, *80*, 101440.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baddeley, A. (2002). Is working memory still working? *European Psychologist*, *7*, 85–97.

Baddeley, A., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Academic press.

Blankenship, T. L., Slough, M. A., Calkins, S. D., Deater-Deckard, K., Kim-Spoon, J., & Bell, M. A. (2019). Attention and executive functioning in infancy: Links to childhood executive function and reading achievement. *Developmental Science*, *22*, e12824.

Bock, J. K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45–93.

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, *42*, 563–570.

Burgoyne, A., & Engle, R. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science*, *29*(6), 624–630.

Burgoyne, A. P., Hambrick, D. Z., & Altmann, E. M. (2019a). Is working memory capacity a causal factor in fluid intelligence? *Psychonomic Bulletin & Review*, *26*, 1333–1339.

Burgoyne, A. P., Hambrick, D. Z., & Altmann, E. M. (2019b). Placekeeping ability as a component of fluid intelligence: Not just working memory capacity. *The American Journal of Psychology*, *132*, 439–449.

Burgoyne, A. P., Tsukahara, J. S., Draheim, C., & Engle, R. W. (2020). Differential and experimental approaches to studying intelligence in humans and nonhuman animals. *Learning & Motivation*, *72*.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 77–126.

Chiou, J. S., & Spreng, R. A. (1996). The reliability of difference scores: A re-examination. *Journal of Consumer Satisfaction Dissatisfaction and Complaining Behavior*, *9*, 158–167.

Conway, A. R., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*, 331–335.

Cronbach, L. J., & Furby, L. (1970). How should we measure "change" – or should we? *Psychological Bulletin*, *74*, 68–80.

Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica*, *50*, 291–323.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Memory and Language*, *19*, 450–466.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433.

Delaney, P. F. (2018). The role of long-term working memory and template theory in contemporary expertise research. *Journal of Expertise*, *3*, 155–161.

Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What item response theory can tell us about the complex span tasks. *Psychological Assessment*, *30*, 116–129.

Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task-switching as a case example. *Perspectives on Psychological Science*, *11*, 133–155.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and development research: A review and commentary on problems and alternatives. *Psychological Bulletin*, *145*, 508–535.

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, *150*(2), 242-275.

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *The Journal of Educational Research*, *84*, 253–262.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

Engel de Abreu, P. M. J. E., Gathercole, S. E., & Martin, R. (2011). Disentangling the relationship between working memory and language: The roles of short-term storage and cognitive control. *Learning and Individual Differences*, *21*, 569–574.

Farnham-Diggory, S., & Gregg, L. W. (1975). Short-term memory function in young readers. *Journal of Experimental Child Psychology*, *19*, 279–298.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*, 226–236.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*, 101–135.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Erlbaum.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Psychology Press.

Hale, S., Myerson, J., Rhee, S. H., Weiss, C. S., & Abrams, R. A. (1996). Selective interference with the maintenance of location information in working memory. *Neuropsychology*, *10*, 228–240.

Hambrick, D. Z., & Altmann, E. M. (2015). The role of placekeeping ability in fluid intelligence. *Psychonomic Bulletin & Review*, *22*, 1104–1110.

Harrison, T. L. (2017). *N-back as a measure of working memory capacity*. (Ph.D. dissertation, Georgia Institute of Technology). http://hdl.handle.net/1853/58762

Hedge, C., Powell, G., Bompas, A., & Sumner, P. (2020, February 1). Strategy and processing speed eclipse individual differences in control ability in conflict tasks. PsyArXiv. https://doi.org/10.31234/osf.io/vgpxq

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166–1186.

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, *136*, 217–240.

Hunt E, Frost N, Lunneborg C. (1973). Individual differences in cognition: A new approach to intelligence. In G. H. Bower, (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 87–122). Academic Press

Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General*, *108*, 151–181.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

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

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*, 637–671.

Kane, M. J., & Engle, R. W. (2003). 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*, *132*, 47–70.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford.

Kramer, A. F., Humphrey, D. G., Larish, J. F., & Logan, G. D. (1994). Aging and inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging*, *9*, 491–512.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203

Martin, J. D., Shipstead, Z., Harrison, T. L., Redick, T. S., Bunting, M., & Engle, R. W. (2020). The role of maintenance and disengagement in predicting reading comprehension and vocabulary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 140–154.

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, *141*, 302–320.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 56–60.

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain–general measure of working memory capacity. *Behavior Research Methods*, *47*, 1343–1355.

Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, *274*, 81–93.

Perfetti, C. A., & Goldman, S. R. (1976). Discourse memory and reading comprehension skill. *Journal of Verbal Learning and Verbal Behavior*, *15*, 33–42.

Perfetti, C. A., & Lesgold, A. M. (1977). Discourse comprehension and sources of individual differences. In M. A. Just and P. A. Carpenter (Eds.), *Cognitive processes in comprehension*. Lawrence Erlbaum Associates.

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales*. Oxford Pyschologists Press.

Redick, T. S., & Lindsey, D. R. (2013). Complex span and *n*-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *20*, 1102–1113.

Rizzo, N. D. (1939). Studies in visual and auditory memory span with special reference to reading disability. *The Journal of Experimental Education*, *8*, 208–244.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, 26, 772–789.

Rouder, J., Kumar, A., & Haaf, J. M. (2019, March 25). Why most studies of individual differences with inhibition tasks are bound to fail. https://doi.org/10.31234/osf.io/3cjr5

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*, 609–612.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*, 4–27.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, *11*, 771–799.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Swanson, H. L., & Ashbaker, M. H. (2000). Working memory, short-term memory, speech rate, word recognition and reading comprehension in learning disabled readers: Does the executive system have a role? *Intelligence*, *28*, 1–30.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, *1*, 270–275.

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

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635–654.

Waters, G., Caplan, D. & Hildebrandt, N. (1987) Working memory and written sentence comprehension. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 531–555). Erlbaum.