Journal of Experimental Psychology: General

Are Individual Differences in Attention Control Related to Working Memory Capacity? A Latent Variable Mega-Analysis

Nash Unsworth, Ashley L. Miller, and Matthew K. Robison

Online First Publication, November 19, 2020. http://dx.doi.org/10.1037/xge0001000

CITATION

Unsworth, N., Miller, A. L., & Robison, M. K. (2020, November 19). Are Individual Differences in Attention Control Related to Working Memory Capacity? A Latent Variable Mega-Analysis. *Journal of Experimental Psychology: General*. Advance online publication. http://dx.doi.org/10.1037/xge0001000

Journal of Experimental Psychology: General

© 2020 American Psychological Association
ISSN: 0096-3445

https://doi.org/10.1037/xge0001000

Are Individual Differences in Attention Control Related to Working Memory Capacity? A Latent Variable Mega-Analysis

Nash Unsworth[1], Ashley L. Miller[1], and Matthew K. Robison[2]
[1] Department of Psychology, University of Oregon
[2] Department of Psychology, Arizona State University

The current study examined whether there are coherent individual differences in attention control abilities and whether they are related to variation in working memory capacity. Data were pooled from multiple studies over 12 years of data collection. Mega-analyses on the combined data set suggested that most of the attention control measures had adequate reliabilities and were weakly to moderately related to one another. A number of latent variable mega-analyses suggested that the attention control measures loaded onto a broad attention control factor and this factor was consistently related to working memory capacity. Furthermore, working memory capacity was generally related to each individual attention control measure. These results provide important evidence for the notion that there is a coherent attention control factor and this factor is related to working memory capacity consistent with much prior research.

Keywords: attention control, individual differences, working memory capacity

Supplemental materials: https://doi.org/10.1037/xge0001000.supp

The ability to control our attention to focus on important information and block potential distracting information is critical for a number of tasks and situations we encounter on a daily basis. These range from the relatively mundane such as trying to concentrate during a boring meeting and prevent daydreaming about an upcoming vacation to concentrating on driving during a blizzard while your child is fussy in the backseat. In both cases, attention control processes are needed to maintain attention on task. By attention control we mean the set of processes that allow us to focus selectively and actively maintain task-relevant information in order to guide thought and action in the presence of internally or externally distracting information. Importantly, it is thought that individuals differ greatly in their attention control (AC) abilities. Individuals high in AC are better at controlling aspects of their attention to actively maintain goal-relevant information to successfully perform a task than are individuals low in AC. Furthermore, these differences are especially pronounced under conditions of high interference or distraction in which attentional capture away from task- or goal-relevant information is likely (e.g., Engle & Kane, 2004). Thus, high AC individuals are better at preventing interference or distraction than low AC individuals, and this AC ability is needed in a host of activities regardless of specific stimulus or processing domains. The AC construct is also sometimes referred to as inhibition, interference resolution, executive attention, or executive control and is typically measured with tasks that assess response inhibition and/or interference control (e.g., Chuderski & Jastrzebski, 2018; Engle & Kane, 2004; Friedman & Miyake, 2004; Kane et al., 2016; Karr et al., 2018; Rey-Mermet et al., 2018; Rey-Mermet et al., 2019; Von Gunten et al., 2019) and is likely similar to the Common Executive Function construct noted by Miyake and Friedman (2012). Additionally, as seen below, we and others have included measures of sustained attention into the overall measurement of AC abilities. Thus, the current AC construct is conceptually similar to related constructs.

If individual differences in AC are important, then AC abilities should be related to other important cognitive abilities such as working memory and fluid intelligence. Indeed, one prominent theory of individual differences in working memory capacity suggests that a main contribution to variation in working memory capacity and a major reason that working memory capacity (WMC) is related to fluid intelligence are differences in AC abilities (e.g., Engle, 2002; Engle & Kane, 2004; Kane & Engle, 2002). As such, AC abilities have been hypothesized as being a core cognitive construct that lies at the heart of individual differences in broad abilities. Recently there has been considerable debate as to whether there is evidence for AC abilities as a psychometric construct with some studies suggesting that there is an AC latent factor that is related to other cognitive abilities such as WMC, whereas other studies suggest there is little evidence for AC as a psychometric construct.

Evidence Consistent With Individual Differences in Attention Control

Evidence in support of the claim that there are important individual differences in AC abilities comes from a variety of studies that have examined relations between performance on various AC

Nash Unsworth https://orcid.org/0000-0001-5169-1647

Matthew K. Robison is now at the Department of Psychology, University of Texas at Arlington.

We thank Cameron S. Kay for help with the simulations and the multi-level analyses.

Correspondence concerning this article should be addressed to Nash Unsworth, Department of Psychology, University of Oregon, Eugene, OR 97403, United States. Email: nashu@uoregon.edu

tasks and WMC. For example, individual differences in WMC are related to performance on AC tasks such as dichotic listening (Colflesh & Conway, 2007; Conway et al., 2001), Stroop interference (Kane & Engle, 2003; Hutchison, 2011; Long & Prat, 2002; Meier & Kane, 2013; Morey et al., 2012; Unsworth, Redick, et al., 2012), flanker interference (Heitz & Engle, 2007; Redick & Engle, 2006; Unsworth, Redick, et al., 2012), performance on the antisaccade task (Kane et al., 2001; Unsworth, Schrock, & Engle, 2004), performance on the psychomotor vigilance task (Unsworth, Redick, Lakey, & Young, 2010; Unsworth & Robison, 2020), performance on the Sustained Attention to Response Task (SART; McVay & Kane, 2009), performance on versions of go/no-go tasks (Redick et al., 2011), performance on the AX-CPT task (Redick, 2014; Redick & Engle, 2011; Richmond et al., 2016), performance on cued visual search tasks (Poole & Kane, 2009), and performance on some versions of the Simon task (Meier & Kane, 2015; Weldon et al., 2013). Additional research comes from a number of latent variable studies that have demonstrated that AC tasks like antisaccade, Stroop, flankers, psychomotor vigilance, go/no-go, and others are weakly to moderately correlated with one another and tend to load on the same factor in a confirmatory factor analysis (e.g., Chuderski et al., 2012; Chuderski & Jastrzebski, 2018; Draheim et al., 2020; Friedman et al., 2008; Friedman & Miyake, 2004; Gärtner & Strobel, 2019; Himi et al., 2019; James et al., 2018; Kane et al., 2016; MacKillop et al., 2016; McVay & Kane, 2012; Miyake et al., 2000; Paap et al., 2020; Redick et al., 2016; Robison & Unsworth, 2018; Shipstead et al., 2014; Stahl et al., 2014; Unsworth & Spillers, 2010; Unsworth & McMillan, 2014, 2017; Unsworth et al., 2014; Venables et al., 2018; Von Gunten et al., 2019; Was, 2007). This latent AC factor tends to correlate strongly with latent WMC (e.g., Chuderski & Jastrzebski, 2018; Kane et al., 2016; McVay & Kane, 2012; Redick et al., 2016; Shipstead et al., 2014; Unsworth & Spillers, 2010; Unsworth & McMillan, 2014, 2017; Unsworth et al., 2014), fluid intelligence (Chuderski & Jastrzebski, 2018; Redick et al., 2016; Shipstead et al., 2014; Unsworth & Spillers, 2010; Unsworth & McMillan, 2014, 2017; Unsworth et al., 2014), and long-term memory factors (Shipstead et al., 2014; Unsworth, 2019; Unsworth & Spillers, 2010; Unsworth et al., 2014). For example, Unsworth and Spillers (2010) had participants perform a number of AC tasks (antisaccade, color-word Stroop, arrow flankers, and psychomotor vigilance) along with measures of WMC, fluid intelligence, and long-term memory. They found that all of the AC tasks correlated and loaded on the same factor. Importantly, this latent AC factor was correlated with WMC (.58), fluid intelligence (.45), and long-term memory (.60). Similarly, Unsworth and McMillan (2014) had participants perform five AC tasks along with measures of WMC and fluid intelligence. Unsworth and McMillan found that all of the AC tasks loaded onto a latent AC factor and this factor was correlated with WMC (.62) and fluid intelligence (.78). Similar results were found by Redick et al. (2016) who administered six AC tasks along with multiple measures of WMC and fluid intelligence. Reanalyzing their data, Unsworth et al. (2015) found that all of the AC tasks loaded onto an AC latent factor and this factor was correlated with both WMC (.76) and fluid intelligence (.75). Thus, there seems to be considerable evidence for the notion that there are coherent individual differences in AC abilities and these AC abilities are related to other cognitive abilities including WMC.

Recent research has also suggested that AC abilities are related to self-reports of off-task thinking during the AC tasks (Kane et al., 2016; McVay & Kane, 2012; Robison & Unsworth, 2018; Unsworth & McMillan, 2014, 2017). That is, those participants who report more mind-wandering and external distraction during the AC (and other) tasks tend to perform worse on those very same tasks. This suggests that the ability to control attention to task-related distractors (i.e., flashing cues in the antisaccade, distractor items in flankers) shares considerable variance with the ability to control attention to task-irrelevant distractors (i.e., mind-wandering about a fight with your spouse, distraction from a flickering overhead light in the run room). Furthermore, individual differences in AC abilities assessed in the laboratory even seem to predict real-world attention failures in some situations (Kane et al., 2017; Unsworth, McMillan, et al., 2012; Unsworth & McMillan, 2017). For example, Unsworth, McMillan, et al. (2012) had participants perform a number of tasks in the laboratory (AC, WMC, prospective memory, retrospective memory) and then carry a diary around for a week logging their various cognitive failures. We found that the most common everyday attention *failures* (distracted during class, mind-wandering during class; distracted during study) loaded onto a common latent factor and this factor was correlated with AC abilities assessed in the laboratory (−.53) and WMC (−.46). These results provide important ecological validity for individual differences in AC abilities.

In the studies noted above, AC was treated as a single unitary construct, but several studies have suggested that AC abilities might be better conceived as several distinct, yet interrelated abilities. For example, Friedman and Miyake (2004; see also Pettigrew & Martin, 2014; Stahl et al., 2014) examined the notion that prepotent response inhibition (measured with tasks like antisaccade and Stroop) was distinct from resistance to distractor interference (measured with tasks like flankers). They found that the tasks were weakly related and loaded onto two distinct factors. Furthermore, they found that the prepotent response inhibition and resistance to distractor interference factors were correlated (.67), suggesting that they shared considerable variance, but were likely distinct. More recently, Kane et al. (2016) tested a similar factor structure to assess differences between the ability to *restrain attention* and prevent prepotent responses from guiding behavior (e.g., preventing the flashing cue in the antisaccade task from capturing attention) and the ability to *constrain attention* to target items among distractors (e.g., to zoom attention in on target items in the flanker task). Participants performed multiple restraint and constraint tasks along with measures of working memory capacity. Kane et al. found that restraint and constraint could be modeled as two separate factors that were strongly correlated (.60) and both were related to WMC (restraint = −.64; constraint = −.40) and off-task thinking (restraint = .37; constraint = .33). Reanalyses of Redick et al. (2016) also suggests the presence of both restraint and constraint factors that are correlated with each other (.74) and with WMC (restraint = −.89; constraint = −.53). Thus, despite limited research, several studies suggest that AC abilities can be broken down into distinct restraint and constraint abilities. Additional research has suggested the possibility of a third subcomponent of AC abilities in terms of the ability to *sustain attention* across both short and long intervals and prevent lapses of attention (Kane et al., 2016; Unsworth, 2015; Unsworth, Spillers, et al., 2009; Unsworth et al., 2010; Unsworth & Robison, 2020; Unsworth &

Spillers, 2010; Unsworth et al., 2020). This ability is seen as important even in situations and tasks were there are really no strong task-relevant distractors (i.e., no flashing cues, no flankers), but where it is critical to keep attention focused on the current task to prevent off-task distractors (mind-wandering) from hijacking attention away. Prior research has shown that measures of restraining, constraining, and sustaining attention load onto the same AC factor that is related to other factors (Unsworth, Spillers, et al., 2009; Unsworth & Spillers, 2010; Unsworth & McMillan, 2014; Unsworth, McMillan, et al., 2012). Additionally, Unsworth and Robison (2017b) found that it was possible to extract a higher-order AC factor based on lower-order restraint/constraint, sustain, and off-task thinking factors and this higher-order AC factor was related to WMC. Thus, recent research suggests that AC abilities can be fractionated into distinct, yet correlated abilities and these distinct AC abilities can be accounted for by a higher-order AC factor. Collectively, prior research suggests that there are robust individual differences in AC abilities that are related to other important cognitive abilities such as WMC.

Evidence Inconsistent With Individual Differences in Attention Control

Despite considerable evidence for individual differences in AC and their relation to WMC, a number of studies have cast doubt on the existence of AC as a construct. Specifically, a number of recent studies have suggested that many AC tasks demonstrate weak and near zero correlations at the task level, resulting in an inability to find a coherent latent AC factor (e.g., De Simoni & von Bastian, 2018; Gärtner & Strobel, 2019; Keye et al., 2009; Krumm et al., 2009; Rey-Mermet et al., 2018, 2019, 2020; Wilhelm et al., 2013; see also Karr et al., 2018). Thus, unlike the prior research discussed above, these studies suggest that AC tasks do not correlate and there is not a general AC factor. For example, Rey-Mermet et al. (2018) administered 11 AC tasks to a sample of older and younger adults. Rey-Mermet et al. found that although most of the tasks had decent reliability estimates, the correlations among the tasks were relatively weak with most correlations hovering near zero. Fitting a one factor model in which all of the AC tasks loaded onto a single factor resulted in a good fit to the data with most of the tasks loading (although many weakly) onto the general factor. Fitting a two-factor model in which prepotent response inhibition (restraint) and resistance to distractor interference (constraint) were modeled as separate factors resulted in a better fit to the data and suggested that the two factors were correlated (.64) similar to prior research. Importantly, however, this correlation was not significant, likely due to an incredibly large standard error. Furthermore, computing Bayes Factors suggested ambiguous evidence for whether there was a correlation between the two factors. Given weak task-level correlations and ambiguous evidence for the latent factor structure, Rey-Mermet et al. suggested that there is not a general AC construct, but rather individual differences reflect task-specific abilities. Rey-Mermet et al. (2018) further suggested that prior research which has found a common AC factor were likely influenced by episodic memory and associative learning abilities which were confounded with many of the AC tasks. As such, they suggested that there is weak evidence for AC abilities as a distinct psychometric construct.

In another recent study, Rey-Mermet et al. (2019) again examined whether a common AC factor could be found. In this study, Rey-Mermet et al. suggested that prior research which has found a coherent AC factor which was related to other abilities was likely due to the fact that the AC tasks were confounded with general processing speed (see also Jewsbury et al., 2016). That is, many AC tasks rely on reaction time (RT) as the main dependent measure (or differences in RT), and thus individual differences in performance may be simply due to differences in processing speed. Indeed, several prior studies have found strong relations between AC factors and processing speed factors (Friedman et al., 2008; James et al., 2018; Salthouse, 2005; Salthouse et al., 2003) and some studies have had to merge the AC and processing speed tasks into a single factor (Hedden & Yoon, 2006). Thus, it seems possible that some prior studies were not assessing AC abilities, but were actually assessing processing speed. To examine this, Rey-Mermet et al. (2019) used a response-deadline procedure to calibrate the time to respond for each participant in each task in order to potentially control for processing speed differences, speed–accuracy trade-offs, and to move AC variance into accuracy. Participants performed seven AC tasks along with measures of WMC and fluid intelligence. Although these new accuracy-based AC tasks demonstrated good reliability, the correlations among the measures were uniformly weak and near zero. Similar to their prior research, Rey-Mermet et al. (2019) were unable to find a coherent AC factor. Furthermore, examining relations between WMC and fluid intelligence with each AC task suggested that none of the relations were significant. Rey-Mermet et al. (2019) suggested that the results challenge the existence of AC as a psychometric construct and challenge the notion that AC abilities are related to WMC and fluid intelligence. Collectively, a number of recent studies have been unable to find evidence for a coherent AC latent variable and this has led some to question whether it is feasible to consider AC as a psychometric construct (Schubert & Rey-Mermet, 2019).

The Present Study

Given the recent debate on the nature of individual differences in AC, in the current study we sought to examine this issue by conducting a number of mega-analyses. Whereas in a traditional meta-analysis effects from prior studies are synthesized, in mega-analysis (also known as integrative data analysis; Curran & Hussong, 2009; or meta-analysis with individual data; Blettner et al., 1999) individual raw data are pooled across multiple studies and analyzed (see e.g., Bialystok et al., 2010; Costafreda, 2009; Curran et al., 2018; Hussong et al., 2008; Robison & Unsworth, 2016; Scoboria et al., 2017). As such, mega-analyses can drastically increase overall power and can result in more precise estimates of effects. To conduct the mega-analysis, we combined data collected in our laboratory (from both the University of Georgia and the University of Oregon) over the last 12 years from several prior studies where participants completed a number of AC and WMC tasks. These studies include Unsworth, Spillers, and Brewer (2009); Unsworth, Miller, et al., (2009); Unsworth and Spillers (2010); Brewer and Unsworth (2012); Unsworth, Brewer, and Spillers (2012); Unsworth and McMillan (2014); Unsworth and McMillan (2017); Robison and Unsworth (2017a; Experiment 1), Robison and Unsworth (2018), Robison et al. (in press); Unsworth, Robison, and Miller (2019), as well as data that have not yet been published. These data reflect all the relevant data available from these labs and were not selected for their prior demonstrations of AC latent factors. Excluded data came from studies where the timing of the AC tasks was changed to accommodate pupillometry (e.g., Unsworth & Robison, 2017a; Unsworth et al., 2020) or where experimental manipulations were done on the tasks (e.g., Unsworth & Robison, 2020). We also excluded data where some of the data were collected in our laboratory and some of the data were collected in other laboratories (e.g., Redick et al., 2016).

In each study, participants performed various AC tasks including antisaccade, color-word Stroop, arrow flankers, psychomotor vigilance, and the sustained attention to response task (SART) along with measures of WMC (operation span, symmetry span, and reading span). We have specifically used these tasks in prior research because they are thought to measure restrain, constrain, and sustain components of a broader AC ability (e.g., Unsworth, Spillers, et al., 2009; Unsworth & Spillers, 2010; Unsworth et al., 2012). That is, measures from antisaccade, Stroop, and no-go accuracy in the SART are thought to partially index the ability to restrain attention and override a prepotent response. Measures from flankers are thought to partially index the ability to constrain attention to target information in the presence of distractors. Whereas measures from the psychomotor vigilance and RT variability in the SART are thought to partially index the ability to sustain attention and prevent lapses of attention. Furthermore, in prior research we have been careful to use a mix of dependent measures including accuracy, RT difference scores, as well as variability in RT in order to ensure that the results are not simply due to using the same dependent measures (e.g., all RTs) as well as avoiding using only RT difference scores which can have poor reliability and other psychometric properties. Because each study utilized nearly identical versions of tasks and similar samples, we reasoned that pooling data across the studies into a combined dataset would provide a powerful test of AC as a psychometric construct.

With the combined dataset we addressed a number of specific questions. First, we asked are AC tasks reliable. Mixed evidence for AC abilities could be partially due to the fact that AC measures can be unreliable and are multidimensional with multiple processes contributing to performance (task impurity). Although some studies have demonstrated that some AC tasks have good reliability (such as antisaccade), other tasks that typically rely on difference scores (such as flankers and Stroop) typically have poorer reliability estimates (e.g., Friedman & Miyake, 2004; Hedge et al., 2018; Kane et al., 2016; Paap & Sawi, 2016; Redick et al., 2016; Rey-Mermet et al., 2018; Unsworth & McMillan, 2014). Thus, one reason that many AC tasks demonstrate weak correlations with each other is because some of the measures are simply not very reliable. Low reliability of AC and other executive control tasks has long been recognized as a major problem (Burgess, 1997; Rabbitt, 1997), and recent investigations similarly suggest unreliability as a key issue for several AC tasks. Indeed, issues with reliability and high measurement error have led some researchers to conclude that individual differences studies using AC tasks are bound to fail (Rouder, Kumar, & Haaf, 2019). Thus, we first examined the extent to which various AC tasks are reliable in the large combined dataset.

Second, we asked whether different AC tasks are related to one another. That is, are various AC tasks related to one another at the task level and related sufficiently to form a coherent AC factor? As noted previously, some prior research has found that various AC tasks are weakly to moderately related at the task level and there is sufficient systematic variance across tasks to form an AC latent factor. Other research, however, has suggested that many AC tasks demonstrate almost no relation at the task level, resulting in an inability to form a latent AC factor. Thus, a key question is the extent to which various AC tasks are related to one another and whether they are suitably related to form a latent factor.

Finally, we asked whether AC is related to WMC. That is, if there is a coherent AC latent factor, is this factor related to individual differences in WMC? Furthermore, is WMC related to individual differences on each AC task separately? As reviewed above, a number of studies have suggested that WMC (based on either a single WMC task or a composite of several tasks) is related to performance on each of the AC tasks used. Furthermore, a number of studies have found that AC and WMC tend to correlate at the latent level. Given the prominence of AC in various theories of WMC, it is expected these two constructs should be related at both the task and latent levels. However, recent research has cast doubt on these findings as Rey-Mermet et al. (2019) found that WMC was not related to any of the AC measures in their study. As such, Rey-Mermet et al. concluded that AC abilities were unlikely to be related to WMC.

By pooling data across a number of prior studies that have utilized similar tasks, the current study is in a unique position to answer important questions on the nature of individual differences in AC. These issues are critically important given the prominence of AC as an explanatory construct in a number of domains. Specifically, AC is thought to be a critical theoretical construct in terms of models that explain individual variation in WMC and the ability of WMC to predict higher-level constructs like fluid intelligence (Engle & Kane, 2004; Unsworth & Engle, 2007). AC is also theorized to be important in explaining age differences (Hasher & Zacks, 1988), neuropsychological differences (van Zomeren & Brouwer, 1994) and is thought to be a key component of intelligence (Duncan et al., 1996; Kane & Engle, 2002; Schneider & McGrew, 2012; Unsworth et al., 2014). But, if there is no evidence for AC as a psychometric construct, then this provides a serious challenge to all models that rely on AC as an explanatory construct. As such there is an important need to examine the validity of AC as a psychometric construct.

Method

Participants

Data were pooled across multiple studies conducted in our laboratory at the University of Georgia (age M = 19.19, SD = 1.65; 65.1% Female) and the University of Oregon (age M = 19.48, SD = 1.90; 61.8% Female) over the last 12 years. The studies were approved by the Institutional Review Boards at the University of Georgia and the University of Oregon. The overall number of participants was $N = 3082$. For each task there were different numbers of participants with available data (see Table 1). In the combined dataset we examined accuracy and RTs for the different measures. Data from participants who had mean RTs < 150 ms and mean RTs $> 5{,}000$ ms were post hoc excluded after examining the pooled data. This resulted in data for six participants on the Stroop and four participants on flanker being excluded because of long RTs. We also excluded participants who had accuracy lower than 50% on congruent and/or neutral trials on the Stroop and flanker. This resulted in data for 11 participants being excluded on Stroop data for 22 participants being excluded on flanker.[1]

Procedure

After signing informed consent, participants completed various combinations of operation span, symmetry span, reading span, antisaccade, flankers, Stroop, psychomotor vigilance task, and SART. In several studies, additional tasks (such as fluid intelligence and long-term memory) were also given, but are not included in the mega-analysis.

Attention Control (AC) Tasks

Antisaccade

In this task (Kane et al., 2001) participants were instructed to stare at a fixation point which was onscreen for a variable amount of time (200–2,200 ms). A flashing white "=" was then flashed 12.7 cm either to the left or right of fixation for 100 ms. The target stimulus (a B, P, or R) then appeared onscreen for 100 ms, followed by masking stimuli (an H for 50 ms followed by an 8, which remained onscreen until a response was given). The participants' task was to identify the target letter by pressing a key for B, P, or R (the keys 4, 5, 6 on the numberpad) as quickly and accurately as possible. In the prosaccade condition the flashing cue (=) and the target appeared in the same location. In the antisaccade condition the target appeared in the opposite location as the flashing cue. Participants first completed practice trials to learn the response mapping then completed prosaccade practice trials, and finally completed antisaccade trials. The number of antisaccade trials was slightly different across studies, with some studies having 40 trials, some studies having 50 trials, and other studies having 60 trials. The dependent variable was proportion correct on the antisaccade trials.

Arrow Flankers

This is a variant of the executive control measure in the Attention Network Test (Fan et al., 2002). This version was initially used by Redick and Engle (2006). Participants were presented with a fixation point for 400 ms. This was followed by an arrow directly above the fixation point for 1,700 ms. The participants' task was to indicate the direction the arrow was pointing (pressing the F for left pointing arrows and pressing J for right pointing arrows) as quickly and accurately as possible. On neutral trials the arrow was flanked by two horizontal lines on each side. On congruent trials the arrow was flanked by two arrows pointing in the same direction as the target arrow on each side. Finally, on incongruent trials the target arrow was flanked by two arrows pointing in the opposite direction as the target arrow on each side. All trial types were randomly intermixed. Participants first performed several practice trials and then the real trials. The number of real trials was slightly different across studies with some studies having 90 trials and other studies having 150 trials. The number of congruent, incongruent, and neutral trials was always equal within a study. The main dependent variables were the RT difference between accurate incongruent and neutral trials as well as proportion correct on incongruent trials. We also computed a composite flanker variable that combined the RT difference score and incongruent accuracy into a single measure. Specifically, we z-scored both the RT difference score and incongruent error rates. We then averaged the *z*-scores together. This is a variant of the balanced integration score (Liesefeld et al., 2015; Liesefeld & Janczyk, 2019) and provides a simple means of combining these two putative measures of conflict into a single measure. There are likely other ways of combining measures into a single composite (see Draheim et al., 2019; Liesefeld & Janczyk, 2019 for discussions).

Stroop

Participants were presented with a color word (red, green, or blue) presented in one of three different font colors (red, green, or blue; Stroop, 1935). The participants' task was to indicate the font color via key press (red = 1, green = 2, blue = 3). Participants were told to press the corresponding key as quickly and accurately as possible. Participants first completed a response mapping practice and then practiced the real task. On the real trials, 67% were congruent such that the word and the font color matched (i.e., red printed in red) and the other 33% were incongruent (i.e., red printed in green). The number of trials was slightly different across studies with some studies having 75 trials, some studies having 100 trials, some studies having 120 trials, and other studies having 135 trials. The dependent variables were the difference in mean RT for accurate incongruent and congruent trials as well as proportion correct on the incongruent trials. We also computed a composite Stroop variable that combined the RT difference score and incongruent accuracy into a single measure. Specifically, we *z*-scored both the RT difference score and incongruent error rates. We then averaged the *z* scores together. This is a variant of the balanced integration score (Liesefeld et al., 2015; Liesefeld & Janczyk, 2019) and provides a simple means of combining these two putative measures of conflict into a single measure. There are

[1] Note participants with low accuracy on flanker incongruent trials but high accuracy on congruent and neutral trials were left in the analysis. Excluding these participants led to similar results.

Table 1
Descriptive Statistics for All Measures

Measure	*M*	*SD*	Range	Skew	Kurtosis	Reliability	*N*
Ospan	.77	.17	.96	−1.13	1.53	.74	3057
Symspan	.70	.18	1.00	−.65	.15	.67	2691
Rspan	.74	.18	1.00	−.89	.69	.79	2910
Anti	.53	.15	.80	.27	−.65	.82	3012
Stroop	148.48	97.23	1101.75	.96	2.95	.52	1768
StroopIAcc	.93	.07	.80	−2.77	16.66	.60	1768
Flanker	113.84	68.22	657.39	1.12	2.95	.61	1525
FlankerIAcc	.92	.16	1.00	−3.99	16.99	.91	1541
PVT	539.64	250.22	3367.34	5.77	48.84	.93	2520
SARTsd	160.96	55.44	446.21	1.65	4.69	.85	812
SARTAcc	.43	.21	.94	.14	−.79	.83	812
StroopComp	.00	.78	10.16	2.05	12.34	.60	1768
FlankerComp	−.03	.71	6.53	2.32	8.54	.77	1525

Note. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; Stroop = RT Stroop effect; StroopIAcc = accuracy on incongruent trials in Stroop; Flanker = RT flanker effect; FlankerIAcc = accuracy on incongruent trials in Flanker; PVT = psychomotor vigilance task; SARTsd = standard deviation of reaction times in sustained attention to response task; SARTacc = accuracy on sustained attention to response task; StroopComp = composite variable combining reaction time (RT) Stroop effect with incongruent accuracy; FlankerComp = composite variable combining RT Flanker effect with incongruent accuracy.

likely other ways of combining measures into a single composite (see Draheim et al., 2019; Liesefeld & Janczyk, 2019 for discussions).

Psychomotor Vigilance Task

In the psychomotor vigilance task (PVT) participants were presented with a row of zeros on screen. After a variable amount of time the zeros began to count up in 17-ms intervals from 0 ms (as determined by the 60 Hz monitor refresh rate). The participants' task was to press the spacebar as quickly as possible once the numbers started counting up. After pressing the space bar the response time was left on screen for 1 s to provide feedback to the participants. Interstimulus intervals were randomly distributed and ranged from 1 s to 10 s. The entire task lasted for 10 min for each individual (roughly 75 total trials). The dependent variable was the average RT for the slowest 20% of trials (Dinges & Powell, 1985; Unsworth et al., 2010).

Sustained Attention to Response Task

Participants completed a version of a Sustained Attention to Response Task (SART) with semantic stimuli adapted from McVay and Kane (2009). The SART is a go/no-go task where subjects must respond quickly with a key press to all presented stimuli except infrequent (11%) target trials. In this version of SART, word stimuli were presented in Courier New font size 18 for 300 ms followed by a 900-ms mask. Most of the stimuli (nontargets) were members of one category (animals) and infrequent targets were members of a different category (foods). The number of real trials was slightly different across studies with some studies having 315 trials and other studies having 470 trials. The dependent variables were accuracy for targets and each individual's standard deviation of RT for go trials.

WMC Tasks

Operation Span

Participants solved a series of math operations while trying to remember a set of unrelated letters (F, H, J, K, L, N, P, Q, R, S, T, Y; see Unsworth et al., 2005). Participants were required to solve a math operation, and after solving the operation they were presented with a letter for 1 s. Immediately after the letter was presented the next operation was presented. At recall participants were asked to recall letters from the current set in the correct order by clicking on the appropriate letters. For all of the span measures, items were scored correct if the item was recalled correctly from the current list. Participants were given practice on the operations and letter recall tasks only, as well as two practice lists of the complex, combined task. List length varied randomly from three to seven items. The total possible correct was slightly different across studies with some studies having a maximum score of 50 and other studies having a maximum score of 75. The dependent variable was proportion of items recalled in the correct serial position.

Symmetry Span

Participants recalled sequences of red squares within a matrix while performing a symmetry-judgment task (see Unsworth, Redick, et al., 2009). In the symmetry-judgment task, participants were shown an 8 × 8 matrix with some squares filled in black. Participants decided whether the design was symmetrical about its vertical axis. The pattern was symmetrical half of the time. Immediately after determining whether the pattern was symmetrical, participants were presented with a 4 × 4 matrix with one of the cells filled in red for 650 ms. At recall, participants recalled the sequence of red-square locations by clicking on the cells of an empty matrix. Participants were given practice on the symmetry-judgment and square recall task as well as two practice lists of the combined task. List length varied randomly from two to five items. The total possible correct was slightly different across studies with some studies having a maximum score of 28 and other studies having a maximum score of 42. The dependent variable was proportion of items recalled in the correct serial position.

Reading Span

While trying to remember an unrelated set of letters (F, H, J, K, L, N, P, Q, R, S, T, Y), participants were required to read a sentence and indicated whether or not it made sense (see Unsworth, Redick, et al., 2009). Half of the sentences made sense, whereas the other half did not. Nonsense sentences were created by changing one word in an otherwise normal sentence. After participants gave their response, they were presented with a letter for 1 s. At recall, participants were asked to recall letters from the current set in the correct order by clicking on the appropriate letters. Participants were given practice on the sentence judgment task and the letter recall task, as well as two practice lists of the combined task. List length varied randomly from three to seven items. The total possible correct was slightly different across studies with some studies having a maximum score of 50 and other studies having a maximum score of 75. The dependent variable was proportion of items recalled in the correct serial position.

Results

Descriptive Statistics and Bivariate Correlations

Descriptive statistics for all of the measures are shown in Table 1. As can be seen, the measures had generally acceptable values of reliability (except for the Stroop RT difference score).[2] All reliability estimates are split-half reliabilities. Additionally, most of the measures were approximately normally distributed (except for the psychomotor vigilance task and the composite variables which were positively skewed, and the incongruent accuracy variables which tended to be negatively skewed).[3] Shown in Figure 1 are frequency histograms for both Stroop and flanker RT difference score effects. As can be seen, there seemed to be quite a bit of variability in both effects, which is inconsistent with some recent claims suggesting that there is insufficient between participant variability to find correlational effects (e.g., Hedge et al., 2018).

Correlations, shown in Table 2, were weak to moderate in magnitude (see the online supplemental materials for scatter plots). The three WMC measures were strongly interrelated and demonstrated weaker relations with the AC tasks. Correlations among the AC tasks were weak to moderate in magnitude (i.e., $r = .10$ small, $r = .20$ medium, $r = .30$ large; see Funder & Ozer, 2019; Gignac & Szodorai, 2016) with an average absolute correlation of $r = .15$. Most of the weak relations were between the two tasks utilizing RT difference scores (Stroop and flanker with each other and with the other variables). The average absolute correlation among the AC tasks without the difference scores was $r = .20$. Incongruent accuracy on Stroop trials also demonstrated generally weak correlations. Examining correlations with the composite Stroop and flanker variables that combine the RT difference scores with incongruent accuracy suggested an average absolute correlation of $r = .19$. Generally similar results were obtained when examining Spearman's rho instead of Pearson correlations. Note that given the SART was used only in a few studies, there are fewer participants for some of the relations with SART. In particular, there were only 227 participants for the correlation between the SART variables and the flanker variables and only 577 participants between the SART variables and the Stroop variables. All other relations are based on Ns > 700. Overall, these results suggest that different AC tasks tend to demonstrate weak to moderate correlations with one another.

Relations Between Each Attention Control Measure With Working Memory Capacity

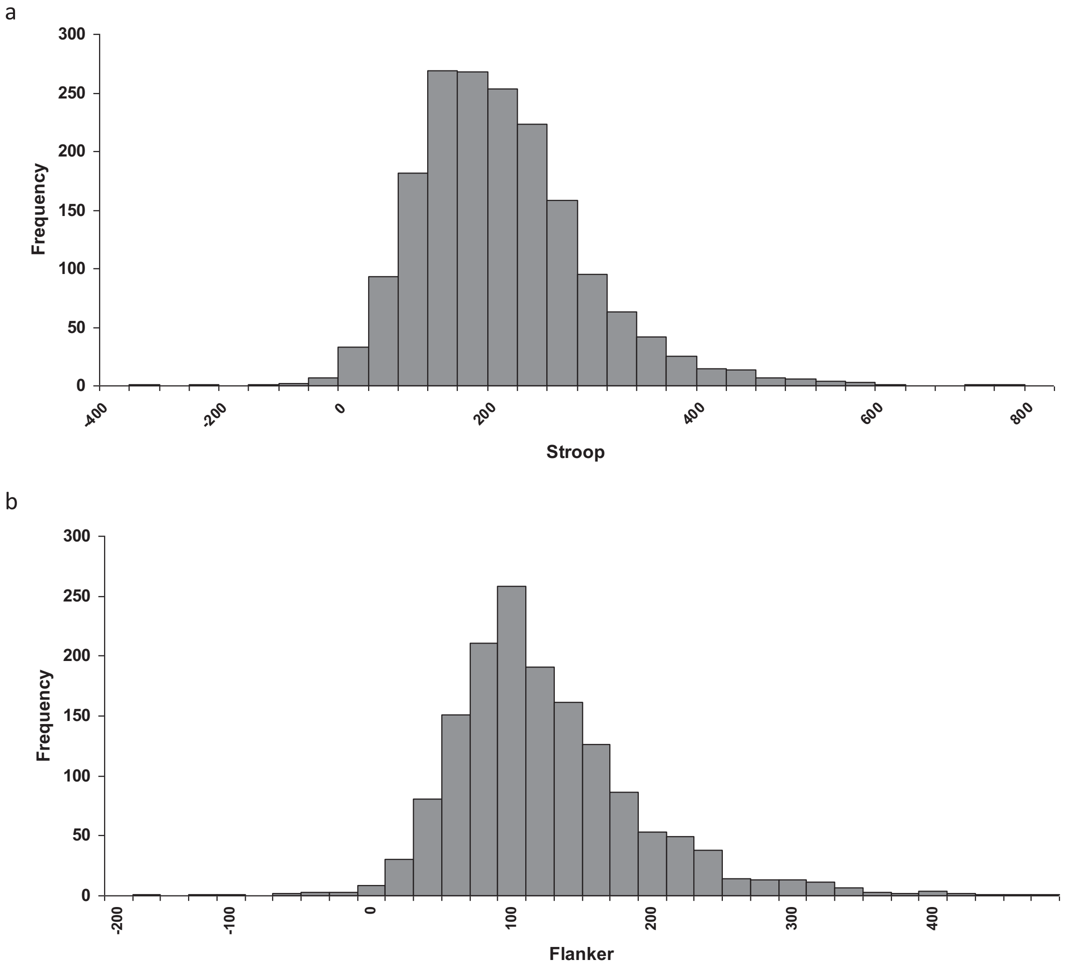
As shown in Table 2 there were weak relations between each AC measure and each WMC measure. Here we examined whether a composite WMC variable would correlate with each AC measure. In many prior studies that have examined the relation between WMC and performance on a single AC task, a composite (typically a z-score or factor composite) has been formed from the various WMC tasks and relations between this composite WMC measure and performance on the AC task are examined. Thus, we wanted to examine whether a composite WMC measure would correlate with each AC measure separately as suggested by prior research. Additionally, as noted previously, Rey-Mermet et al. (2019) found that none of their AC measures were related to WMC, thus we wanted to further examine whether similar results would be obtained in the combined data set. To examine this, we created a z score composite for WMC by first z-scoring each WMC measure and the averaging the resulting z scores. Note, if a participant was missing a WMC measure, the z score was computed based on the WMC measures that were available. Overall similar results were found when using a factor composite for WMC. Shown in Table 3 are the resulting correlations and the N for each bivariate correlation. As can be seen, WMC was weakly to moderately related with each AC measure. All ps < .001, except for the correlation

[2] We also examined test–retest reliability of Stroop in a separate sample of participants. Specifically, 78 participants were tested on the same version of Stroop (67–33 proportion congruency) used in the current study along with a version in which proportion congruency was 50–50. Participants completed the Stroop along with other tasks (such as operation span) and then came back four days later and completed the tasks again (67 participants came back for the second session). For the 67–33 proportion congruency task test–retest reliability for the Stroop effect was $r = .56$, $r = .85$ for accuracy on incongruent trials, and $r = .63$ for the composite variable including the Stroop effect and incongruent accuracy together. For the 50–50 proportion congruency task test–retest reliability for the Stroop effect was $r = .62$, $r = .76$ for accuracy on incongruent trials, and $r = .66$ for the composite variable including the Stroop effect and incongruent accuracy together. Additionally, the Stroop effect in the 67–33 task and the 50–50 tasks were correlated in each session (Session 1 $r = .52$, Session 2 $r = .42$). Operation span demonstrated good test-retest reliability ($r = .76$), consistent with prior research (Unsworth et al., 2005).

[3] Transforming the accuracy variables (with an acrsine transformation) and the psychomotor vigilance task (with a log transformation) resulted in more normally distributed variables but the overall correlations remained the same (see the Appendix for a model based on these transformed variables).

Figure 1

Frequency Distributions for Stroop and Flanker Effects (a) Stroop Effect (b) Flanker Effect



between WMC and Stroop incongruent accuracy (p = .001). Additionally, all Bayes factors > 2,000, except for the correlation between WMC and Stroop incongruent accuracy (BF = 4.76). In particular, the largest numerical relation was with antisaccade, but overall weak to moderate relations were also seen for the Stroop effect, PVT, SART sd, SART accuracy, as well as the Stroop and flanker composite variables. Weaker relations were seen between WMC and incongruent accuracy on both the Stroop and flanker tasks. The average absolute correlation between WMC and each AC task was r = .16. Overall, these results are consistent with much prior research suggesting relations between WMC and performance on each AC task.

Confirmatory Factor Analyses

Next, we used latent variable techniques to test our main questions of interest. For all model testing we used R (R Core Team, 2017) with the *lavaan* package (Rosseel, 2012). Overall similar results were obtained when using Lisrel to fit the models. Due to large amounts of missing data for many of the pairwise relations, we used full information maximum likelihood estimation to utilize all available data points (Enders, 2010). Overall similar results were obtained when examining all available pairwise information (pairwise deletion) and when utilizing multiple imputation. For each model we report several fit statistics. Nonsignificant chi-square tests indicate adequate model fit; with large samples like ours, however, they are nearly always significant. Comparative fit indices (CFI) and Tucker-Lewis indices (TLI) of ≥ .95 indicate adequate fit, whereas the Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) values of ≤.06 indicate good fit (e.g., Schermelleh-Engel et al., 2003).

Relations Between Attention Control and Working Memory Capacity

For our first model we examined whether the various AC tasks would load onto a general AC factor and whether this factor would be related to a WMC factor. Therefore, we specified a model in which eight of the AC measures (antisaccade, Stroop, Stroop incongruent accuracy, flanker, flanker incongruent accuracy, PVT, SART sd, and SART accuracy) were al-

Table 2

Correlations Among the Measures

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Ospan	—												
2. Symspan	**0.44** [.41, .47]	—											
3. Rspan	**0.55** [.52, .57]	**0.37** [.34, .40]	—										
4. Anti	**0.18** [.15, .22]	**0.22** [.19, .26]	**0.2** [.17, .24]	—									
5. Stroop	**−0.16** [−.20, −.11]	**−0.12** [−.16, −.07]	**−0.1** [−.15, −.05]	**−0.13** [−.17, −.08]	—								
6. StroopIAcc	**0.07** [.03, .12]	**0.07** [.02, .11]	*0.04* [−.01, .08]	**0.11** [.06, .16]	**−0.21** [−.25, −.16]	—							
7. Flanker	**−0.11** [−.16, −.06]	**−0.13** [−.18, −.07]	**−0.12** [−.17, −.07]	**−0.18** [−.23, −.13]	0.09 [.02, .17]	*−0.05* [−.13, .02]	—						
8. FlankerIAcc	**0.11** [.06, .16]	0.07 [.01, .13]	**0.13** [.08, .18]	**0.2** [.16, .25]	*−0.02* [−.09, .05]	**0.15** [.08, .22]	**−0.22** [−.27, −.20]	—					
9. PVT	**−0.12** [−.16, −.08]	**−0.18** [−.22, −.13]	**−0.13** [−.17, −.09]	**0.27** [−.31, −.24]	**0.08** [.03, .13]	*−0.03* [−.08, .03]	**0.09** [.03, .15]	**−0.26** [−.32, −.20]	—				
10. SARTsd	**−0.1** [−.17, −.03]	**−0.17** [−.24, −.10]	**−0.18** [−.25, −.11]	**0.28** [−.34, −.21]	*0.05* [−.03, .13]	−0.1 [−.18, −.01]	*0.02* [−.11, .15]	*−0.09* [−.21, .04]	**0.29** [.23, .36]	—			
11. SARTAcc	**0.13** [.06, .19]	**0.17** [.10, .23]	**0.11** [.05, .18]	**0.23** [.16, .30]	*−0.03* [−.11, .05]	**0.26** [.18, .34]	*−0.02* [−.15, .11]	**0.2** [.07, .32]	**−0.2** [−.27, −.13]	**−0.31** [−.37, −.25]	—		
12. StroopComp	**−0.15** [−.19, −.10]	**−0.12** [−.16, −.07]	**−0.09** [−.13, −.04]	**−0.15** [−.20, −.11]	**0.78** [.76, .80]	**−0.78** [−.80, −.76]	0.09 [.02, .17]	**−0.11** [−.18, −.04]	0.07 [.02, .12]	0.09 [.01, .17]	**−0.18** [−.26, −.10]	—	
13. FlankerComp	**−0.15** [−.20, −.10]	**−0.14** [−.20, −.09]	**−0.16** [−.22, −.11]	**−0.25** [−.29, −.20]	*0.07* [−.01, .14]	**−0.11** [−.19, −.04]	**0.83** [.81, .84]	**−0.73** [−.75, −.71]	**0.23** [.17, .29]	*0.06* [−.07, .19]	*−0.13* [−.25, .00]	**0.12** [.05, .19]	—

Note. Bold correlations are significant at $p < .01$; italicized correlations are not significant at $p > .05$; correlations in standard font are correlated at $p < .05$. Values in brackets are 95% confidence intervals. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; Stroop = Stroop effect; StroopIAcc = accuracy on incongruent trials in Stroop; Flanker = flanker effect; FlankerIAcc = accuracy on incongruent trials in Flanker; PVT = psychomotor vigilance task; SARTsd = standard deviation of reaction times in sustained attention to response task; SARTacc = accuracy on sustained attention to response task; StroopComp = composite variable combining RT Stroop effect with incongruent accuracy; FlankerComp = composite variable combining RT Flanker effect with incongruent accuracy.

Table 3

Correlations Between Attention Control Measures and Composite Working Memory Capacity Measure

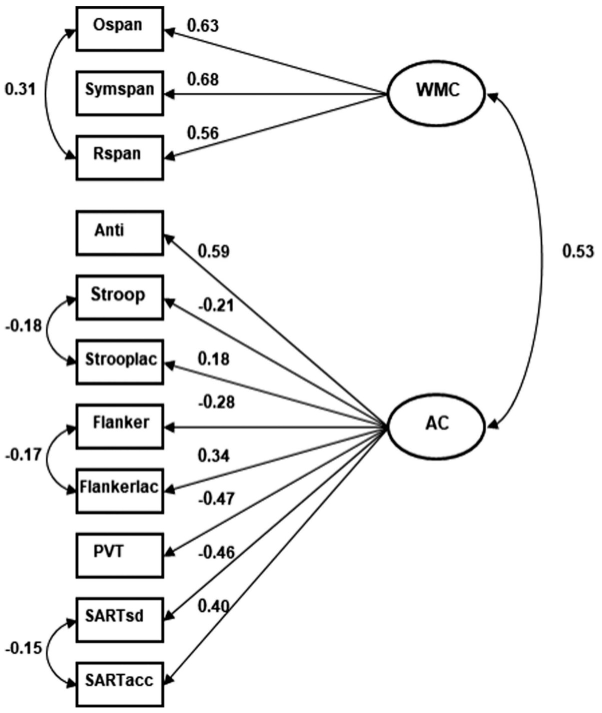
AC measure	Correlation with WMC	95% CI	*N*
Anti	**.25**	[.21, .28]	3,003
Stroop	**−.16**	[−.20, −.11]	1,767
StroopIacc	**.08**	[.03, .12]	1,767
Flanker	**−.14**	[−.19, −.09]	1,516
FlankerIacc	**.12**	[.07, .17]	1,532
PVT	**−.17**	[−.21, −.13]	2,511
SARTsd	**−.19**	[−.26, −.13]	811
SARTacc	**.17**	[.10, .24]	811
StroopComp	**−.15**	[−.20, −.10]	1,767
FlankerComp	**−.18**	[−.23, −.13]	1,516

Note. Bold correlations are significant at $p < .01$. Values in brackets are 95% confidence intervals. Anti = antisaccade; Stroop = Stroop effect; StroopIAcc = accuracy on incongruent trials in Stroop; Flanker = flanker effect; FlankerIAcc = accuracy on incongruent trials in Flanker; PVT = psychomotor vigilance task; SARTsd = standard deviation of reaction times in sustained attention to response task; SARTacc = accuracy on sustained attention to response task; StroopComp = composite variable combining RT Stroop effect with incongruent accuracy; FlankerComp = composite variable combining RT Flanker effect with incongruent accuracy.

lowed to load onto the AC factor. The three WMC measures were allowed to load onto the WMC factor, and the factors were allowed to correlate. We a priori specified residual variances for operation span and reading span to correlate given that these tasks use the same set of stimulus materials and are nearly identical, differing only in the processing task. Additionally, residual variances for variables from the same task for Stroop, flanker, and SART were allowed to correlate. The overall fit of the model was good, $\chi^2(39) = 125.77$, $p < .001$, RMSEA = .03 [.02, .03], CFI = .97, TLI = .96, SRMR = .04. Shown in Figure 2 is the model. As can be seen, all of the AC measures loaded significantly on the AC factor. Most of the loadings were moderate (antisaccade, PVT, SART sd, SART accuracy), but the loadings for Stroop were much weaker. Standard errors of the factor loadings were all .03 except for the SART variables where the standard errors were .04. The overall AC factor was correlated with WMC at .53 (*SE* = .03), consistent with prior research. We also estimated the factor reliability with the Omega coefficient (ω; Raykov, 2001b). Factor reliability was moderate for both factors (WMC ω = .61, AC ω = .54). Note that with correlated errors, estimates of ω tend to be lower (Raykov, 2001a; Savalei & Reise, 2019). Indeed, with the correlated errors taken out of the model, estimates of ω increased (WMC ω = .72, AC ω = .61). We compared this model to a one factor model in which all of the AC and WMC measures were allowed to load onto a single factor. The overall fit of the model was adequate, $\chi^2(40) = 349.70$, $p < .001$, RMSEA = .05 [.045, .055], CFI = .89, TLI = .85, SRMR = .06. Importantly, the one factor model fit significantly worse than the two-factor model, $\Delta\chi^2(1) = 223.93$, $p < .001$. Thus, the two-factor model demonstrated that the AC tasks are sufficiently related to one another to load on the same factor (although the Stroop variables loaded weakly), and this factor was moderately related to WMC. As such, these results are

Figure 2

Confirmatory Factor Analysis Model for Working Memory Capacity (WMC) and Attention Control (AC) for the Full Combined Dataset



Note. Paths connecting latent variables (circles) to each other represent the correlations between the constructs and the numbers from the latent variables to the manifest variables (squares) represent the loadings of each task onto the latent variable. Solid paths are significant at the $p < .05$ level. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; Stroop = Stroop effect; StroopIac = accuracy on incongruent trials in Stroop; Flanker = flanker effect; FlankerIac = accuracy on incongruent trials in Flanker; PVT = psychomotor vigilance task; SARTsd = standard deviation of reaction times (RTs) in sustained attention to response task; SARTacc = accuracy on sustained attention to response task.

inconsistent with recent claims suggesting that a coherent AC factor cannot be found and that AC tasks are not related to WMC.[4]

Next, we tested a number of additional models with the full dataset to examine whether the factor loadings and relation with WMC were driven by specific measures. Specifically, in our next model we examined whether antisaccade was driving the relations seen in the prior model. Prior research has suggested that because antisaccade tends to load the highest on the AC factor this suggests that the factor is really just an antisaccade factor rather than a general AC factor (Rey-Mermet et al., 2019).[5] If this is the case, then taking antisaccade out of the model should result in an inability to find a coherent AC factor, and any resulting factor should not be related to WMC. To test this model, we specified the same model as above but simply did not include the antisaccade as a measure of AC. The overall fit of the model was good, $\chi^2(30) = 119.99$, $p < .001$, RMSEA = .03 [.025, .037], CFI = .96, TLI = .94, SRMR = .04. Shown in Table 4 are the resulting factor loadings as well as the correlation between the factors. As can be seen, all of the measures loaded significantly on the AC factor ($\omega = .44$) and the loadings for each measure were very similar to the loadings in the model that included antisaccade (seen in Figure 2). Furthermore, the correlation between AC and WMC was similar to the prior model (.51). Thus, these results suggest that although antisaccade had the highest numerical loading on the AC factor in the prior model, the factor was not just an antisaccade factor. Taking antisaccade out of the model resulted in nearly identical results as when it was included in the model.

In the next model we examined whether the SART task was unduly influencing the relations. In particular, as shown in Table 1, because the SART was only included in a few prior studies, there were far less data available for this task (indeed only one study included both SART and flanker) which could have resulted in less robust estimates. To examine whether the SART variables were influencing the factor structure we specified the same model as shown in Figure 2, but now excluded both SART variables. All other measures and relations remained the same. The overall fit of the model was good, $\chi^2(23) = 71.38$, $p < .001$, RMSEA = .03 [.019, .033], CFI = .98, TLI = .97, SRMR = .03. Shown in Table 4 are the resulting factor loadings as well as the correlation between the factors. As can be seen, all of the measures loaded significantly on the AC factor ($\omega = .44$) and the loadings for each measure were very similar to the loadings in the full model that included SART (seen in Figure 2) as was the correlation with WMC. Thus, these results suggest that even though there was quite a bit of missing data for the SART variables, this did not seem to influence the overall factor much.

[4] As far as we know, there is no standard way of defining what a coherent factor is. In the current article, we focused on the overall factor loadings as well as the robustness of the factor when various tasks were excluded or different measures were used. In a recent paper, Rey-Mermet et al. (2020) suggested that a good model should include: "1) the Kaiser-Meyer-Olkin (KMO) index–a measure of whether the correlation matrix is factorable–should be larger than .60; (2) most of the error variances needed to be lower than .90; (3) most of the factor loadings had to be significant and larger than .30; (4) no factor should be dominated by a large loading from one task; (5) the amount of shared variance across tasks—that is, 'factor reliability' as assessed by coefficient ω—had to be high (i.e., about.70)." In our models, KMO was .71, most of the error variances were below .90 (except for some of the Stroop and flanker variables) because the factor loadings were less than .30 (note that Criteria 2 and 3 are redundant), the factor was not dominated by a single task, and overall factor reliability measured by ω was moderate across many models. Overall, based on these criteria it does seem like our AC factor is coherent factor.

[5] Rey-Mermet et al. (2019) suggested that in two of our prior studies (Unsworth & Spillers, 2010; Unsworth & McMillan, 2014) antisaccade had a very high loading, whereas the other tasks had very low loadings, suggesting that the AC factor in these studies was really just an antisaccade factor. However, this is not correct. In both studies although antisaccade had the highest numerical loading, the loading was within a standard error of several of the other AC tasks. Furthermore, we reanalyzed data from each of these studies excluding antisaccade from the AC factor. Excluding antisaccade resulted in nearly identical results in which all of the AC tasks loaded on the AC factor, and the AC factor was correlated with other factors including WMC, fluid intelligence, and long-term memory. Thus, although antisaccade had the highest factor loading, the AC factor was not driven by a single task which is inconsistent with Rey-Mermet et al.'s (2019) claims.

Table 4
Standardized Factor Loadings, Standard Errors, and Correlations Between Constructs for Confirmatory Factor Analyses

Construct/measure	No anti	No SART	No IAcc	No RT dif	All acc	One meas
WMC						
Ospan	.63 (.03)	.63 (.03)	.62 (.03)	.63 (.03)	.64 (.03)	.62 (.03)
Symspan	.69 (.04)	.68 (.03)	.69 (.03)	.69 (.03)	.67 (.03)	.69 (.03)
Rspan	.55 (.03)	.56 (.03)	.55 (.03)	.56 (.03)	.56 (.03)	.55 (.03)
AC						
Anti		.60 (.03)	.59 (.03)	.58 (.03)	.57 (.04)	.60 (.03)
Stroop	−.20 (.04)	−.23 (.03)	−.22 (.03)			−.23 (.03)
StroopIAcc	.18 (.04)	.17 (.03)		.19 (.03)	.24 (.04)	
Flanker	−.24 (.04)	−.29 (.03)	−.28 (.03)			−.28 (.03)
FlankerIAcc	.34 (.04)	.35 (.03)		.35 (.03)	.34 (.04)	
PVT	−.48 (.04)	−.45 (.03)	−.46 (.03)	−.49 (.03)		−.44 (.03)
SARTsd	−.48 (.05)		−.49 (.04)	−.47 (.04)		
SARTAcc	.42 (.05)		.39 (.04)	.42 (.04)	.44 (.05)	.39 (.04)
WMC-AC *r*	.51 (.04)	.53 (.03)	.55 (.03)	.51 (.03)	.53 (.04)	.55 (.03)

Note. Standard errors are in parentheses. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; Stroop = RT Stroop effect; StroopIAcc = accuracy on incongruent trials in Stroop; Flanker = RT flanker effect; FlankerIAcc = accuracy on incongruent trials in Flanker; PVT = psychomotor vigilance task; SARTsd = standard deviation of reaction times in sustained attention to response task; SARTacc = accuracy on sustained attention to response task.

A similar model was specified to see whether including incongruent accuracy on the Stroop and flanker tasks influenced the factor structure. That is, typically only the RT difference score is used as the measure of Stroop and flanker, and thus including incongruent accuracy for these tasks may have influenced the overall factor. To see whether this was the case, we specified the same model as before but now excluded incongruent accuracy for both Stroop and flanker tasks. The overall fit of the model was good, $\chi^2(24) = 53.38$, $p = .001$, RMSEA = .02 [.01, .03], CFI = .99, TLI = .98, SRMR = .03. Shown in Table 4 are the resulting factor loadings as well as the correlation between the factors. As can be seen, all of the measures loaded significantly on the AC factor ($\omega = .55$) and the loadings for each measure were very similar to the loadings in the full model that included incongruent accuracy (seen in Figure 2). The correlation with WMC also remained very similar.

We specified a similar model, but now excluding the RT difference score variables for Stroop and flanker. As noted previously, RT difference scores can have poor reliability which influences their potential correlations with other measures. Indeed, as shown in Table 1, the RT difference scores for Stroop and flanker tended to have the lowest reliability estimates. Thus, we tested a model in which the AC factor was composed of the various AC measures, but did not include the RT difference score measures for Stroop and flanker. The overall fit of the model was good, $\chi^2(24) = 85.65$, $p < .001$, RMSEA = .03 [.022, .036], CFI = .98, TLI = .96, SRMR = .04. Shown in Table 4 are the resulting factor loadings as well as the correlation between the factors. As can be seen, all of the measures loaded significantly on the AC factor ($\omega = .55$), and the loadings for each measure were very similar to the loadings in the full model that included the RT difference scores (seen in Figure 2) as was the correlation with WMC.

In our next model we examined whether we could extract an AC factor from only the accuracy measures and whether this factor would correlate with WMC. In particular, Rey-Mermet et al. (2019) suggested that one issue with prior studies is that there was a mismatch of method variance between AC and WMC in that many AC measures relied on RT whereas the WMC measures relied on accuracy. Thus, we wanted to examine whether we could find an accuracy only AC factor and whether this factor would be related to WMC. Therefore, we specified a model in which accuracy from antisaccade, Stroop, flankers, and the SART all loaded onto the AC factor. The overall fit of the model was good, $\chi^2(12) = 39.81$, $p < .001$, RMSEA = .03[.02, .04], CFI = .99, TLI = .98, SRMR = .03. Shown in Table 4 are the resulting factor loadings as well as the correlation between the factors. As can be seen, all of the measures loaded significantly on the AC factor ($\omega = .43$), and the loadings for each measure were very similar to the loadings in the full model that included the RT measures (seen in Figure 2) as was the correlation with WMC.

Finally, we estimated a model in which only one measure per task was used to examine whether using multiple measures for some of the tasks influenced the results. Thus, in this model AC was composed of antisaccade, the Stroop effect, the flanker effect, PVT, and accuracy on the SART (similar results were obtained using standard deviation of RT on the SART). The overall fit of the model was good, $\chi^2(18) = 35.54$, $p = .008$, RMSEA = .02 [.01, .03], CFI = .99, TLI = .99, SRMR = .02. Shown in Table 4 are the resulting factor loadings as well as the correlation between the factors. As can be seen, all of the measures loaded significantly on the AC factor ($\omega = .49$), and the correlation between AC and WMC was similar to the other models.

Collectively, the results suggest that when examining the full dataset that the AC measures are sufficiently related to form a coherent AC factor, and this factor is correlated with WMC. Furthermore, examining various models in which different measures were excluded resulted in nearly identical results as the full model, suggesting that the AC factor was not dependent on the various measures. As such, these results suggest that there are individual differences in broad AC abilities that are related to WMC.

Models With List-Wise Deletion

Although the results from the full dataset are important, given the large amount of missing data associated with some of the relations, it is reasonable to wonder whether the results are simply a product of using full information maximum likelihood. To examine this, we ran a number of models using list-wise deletion to see if generally similar results are obtained when using list-wise deletion compared to when using all available data. Note, for these models list-wise deletion was done only on the AC tasks such that if a participant was missing data on a WMC measure, but had full information on all AC tasks, they were still included in the analysis. Overall similar results are found when doing strict list-wise deletion on all measures. In our first model we examined relations among antisaccade, flanker, and PVT given that these three tasks are thought to provide measures of restraining attention, constraining attention, and sustaining attention (e.g., Kane et al., 2016; Poole & Kane, 2009; Unsworth, Spillers, et al., 2009; Unsworth & Spillers, 2010). In the model we specified antisaccade, flankers (RT difference score and incongruent accuracy), and PVT to load on the AC factor, while the three working memory measures loaded on the WMC factor. The residual variance between operation span and reading span was allowed to correlate as before. With list-wise deletion there were 1,018 participants available for this model. The overall fit of the model was good, $\chi^2(11) = 26.98$, $p = .005$, RMSEA = .04 [.02, .06], CFI = .98, TLI = .97, SRMR = .03. Shown in Table 5 is the resulting model (labeled AFP). As can be seen, all of the measures loaded significantly on the AC factor ($\omega = .48$), and the AC and WMC factors were correlated (.54). Thus, only examining a subset of tasks and participants for which full data were available suggested that tasks thought to represent different aspects of AC are correlated, load onto the same factor, and this factor is related to WMC.

Similar results were obtained when we examined relations among antisaccade, Stroop, and flanker and their relations with WMC. Several of the prior studies which could not find evidence for a general AC factor primarily relied on "inhibition" tasks. With list-wise deletion on the AC measures there were 714 participants available for this model. The overall fit of the model was good, $\chi^2(16) = 33.13$, $p = .007$, RMSEA = .04 [.02, .06], CFI = .98, TLI = .96, SRMR = .03. Shown in Table 5 is the resulting model (labeled ASF). As can be seen, all of the measures loaded significantly on the AC factor ($\omega = .36$), and the AC and WMC factors were correlated (.46). Similar results were obtained when we examined relations among antisaccade, Stroop, and PVT. With list-wise deletion on the AC measures there were 1,420 participants available for this model. The overall fit of the model was good, $\chi^2(11) = 25.49$, $p = .008$, RMSEA = .03 [.02, .05], CFI = .99, TLI = .98, SRMR = .02. Shown in Table 5 is the resulting model (labeled ASP). As can be seen, all of the measures loaded significantly on the AC factor ($\omega = .36$), and the AC and WMC factors were correlated (.56). Similar to the prior models, these results suggest that there is coherent AC factor and this factor is related to a WMC factor.

Table 5

Standardized Factor Loadings, Standard Errors, and Correlations Between Constructs for Confirmatory Factor Analyses With List-Wise Deletion

Construct/measure	AFP	ASF	ASP
WMC			
Ospan	.69 (.05)	.82 (.07)	.64 (.04)
Symspan	.65 (.05)	.58 (.06)	.71 (.04)
Rspan	.64 (.05)	.79 (.08)	.49 (.04)
AC			
Antisaccade	.50 (.04)	.56 (.06)	.59 (.04)
Stroop		−.18 (.05)	−.24 (.04)
StroopIAcc		.15 (.03)	.17 (.04)
Flanker	−.26 (.04)	−.35 (.06)	
FlankerIacc	.45 (.04)	.39 (.06)	
PVT	−.54 (.04)		−.42 (.04)
WMC-AC *r*	.54 (.05)	.46 (.07)	.56 (.05)

Note. Standard errors are in parentheses. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; Stroop = RT Stroop effect; StroopIAcc = accuracy on incongruent trials in Stroop; Flanker = RT flanker effect; FlankerIAcc = accuracy on incongruent trials in Flanker; PVT = psychomotor vigilance task; AFP = antisaccade, flanker, and psychomotor vigilance task model; ASF = antisaccade, Stroop, and flanker model; ASP = antisaccade, Stroop, and psychomotor vigilance task model.

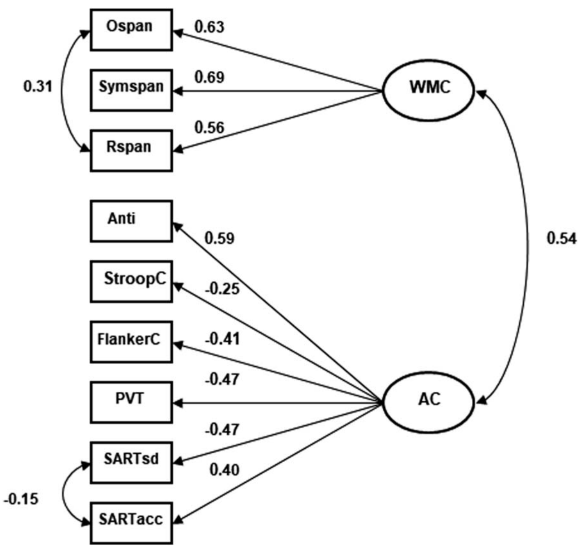
Models Using Flanker and Stroop Composite Variables

For our next set of models, we utilized the composite Stroop and flanker variables that combined the RT difference scores and incongruent accuracy into a single variable to see whether these variables resulted in better overall estimates of AC. In the first model we specified that the six AC measures (antisaccade, Stroop composite, flanker composite, PVT, SART sd, and SART accuracy) loaded onto the AC factor while the three WMC measures loaded onto the WMC factor. The factors were allowed to correlate. The residuals for operation and reading span were also allowed to correlate, and we estimated missing data with full information maximum likelihood. The overall fit of the model was good, $\chi^2(24) = 57.42$, $p < .001$, RMSEA = .02 [.01, .03], CFI = .99, TLI = .98, SRMR = .03. Shown in Figure 3 is the model. As can be seen, all of the measures loaded significantly and moderately on the AC factor (the loadings for Stroop was weaker). Standard errors of the factor loadings were all .03 except for the SART variables where the standard errors were .04. Factor reliability was moderate for both factors (WMC $\omega = .61$, AC $\omega = .59$). Consistent with the prior models, AC and WMC were correlated at .54 (*SE* = .03). Thus, using the composite Stroop and flanker variables resulted in increased loadings for those tasks on the overall AC factor, and this factor was correlated with WMC. Note, if antisaccade is taken out of the model, the resulting model fit the data well, $\chi^2(17) = 51.98$, $p < .001$, RMSEA = .03 [.02, .04], CFI = .98, TLI = .97, SRMR = .03. The factor loadings remained largely the same as those seen in Figure 4, and AC and WMC were correlated (.53), suggesting again that the factor was not driven solely by antisaccade. Similar results were obtained when using the Stroop and flanker composite variables and list-wise deletion (see the Appendix).

Models Including Baseline Reaction Time Measures

Given concerns that many of the AC tasks are influenced by variation in processing speed and do not necessarily reflect variation in AC abilities, we next examined models in which differ-

Figure 3
Confirmatory Factor Analysis for Working Memory Capacity (WMC) and Attention Control (AC) Based on Stroop and Flanker Composite Variables



Note. Paths connecting latent variables (circles) to each other represent the correlations between the constructs and the numbers from the latent variables to the manifest variables (squares) represent the loadings of each task onto the latent variable. Solid paths are significant at the $p < .05$ level. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; StroopC = composite variable combining reaction time (RT) Stroop effect with incongruent accuracy; FlankerC = composite variable combining RT Flanker effect with incongruent accuracy; PVT = psychomotor vigilance task; SARTsd = standard deviation of RTs in sustained attention to response task; SARTacc = accuracy on sustained attention to response task.

ences in baseline RT were incorporated. Specifically, in our first model we ran the same overall model depicted in Figure 2 but now included a Baseline RT factor that was composed of mean correct RTs on congruent trials on the Stroop and Flanker tasks as well as the fastest 20% of RTs on the PVT. We allowed the same residual variances as the full model to correlate as well as residual variances for RT measures from the same task (e.g., Stroop to Stroop congruent RT; flanker to flanker congruent RT, and slowest 20% and fastest 20% in the psychomotor vigilance task). This model should allow for an assessment of relations among WMC, AC, and Baseline RT. The overall fit of the model was good, $\chi^2(67) = 305.04$, $p < .001$, RMSEA = .034 [.03, .04], CFI = .94, TLI = .93, SRMR = .05. Shown in Figure 4a is the model. Standard errors of the factor loadings were all .03 except for the SART variables where the standard errors were .04. Factor reliability was moderate for all factors (WMC ω = .62, AC ω = .50, Baseline RT ω = .66). As can be seen, the AC factor was correlated with WMC at .53 (*SE* = .03) and with the Baseline RT factor at −.86 (*SE* = .03). Furthermore, WMC and the Baseline RT factor were correlated at −.38 (*SE* = .03). Thus, AC and Baseline RT were correlated, but AC demonstrated a stronger correlation with WMC than Baseline RT, suggesting that there was likely important variance shared by AC and WMC that was not shared with Baseline RT. Indeed, constraining the correlations to be equal resulted in a worse model fit, $\Delta\chi^2(1) = 151.51$, $p < .001$, suggesting that AC was more strongly correlated with WMC than Baseline RT was. We also compared this model to a one factor model in which all of the AC and Baseline RT measures were allowed to load onto a single factor. The overall fit of the model was good, $\chi^2(69) = 340.84$, $p < .001$, RMSEA = .036 [.03, .04], CFI = .94, TLI = .92, SRMR = .05. Importantly, the one factor model fit significantly worse than the two-factor model suggesting distinct AC and Baseline RT factors, $\Delta\chi^2(2) = 35.8$, $p < .001$.

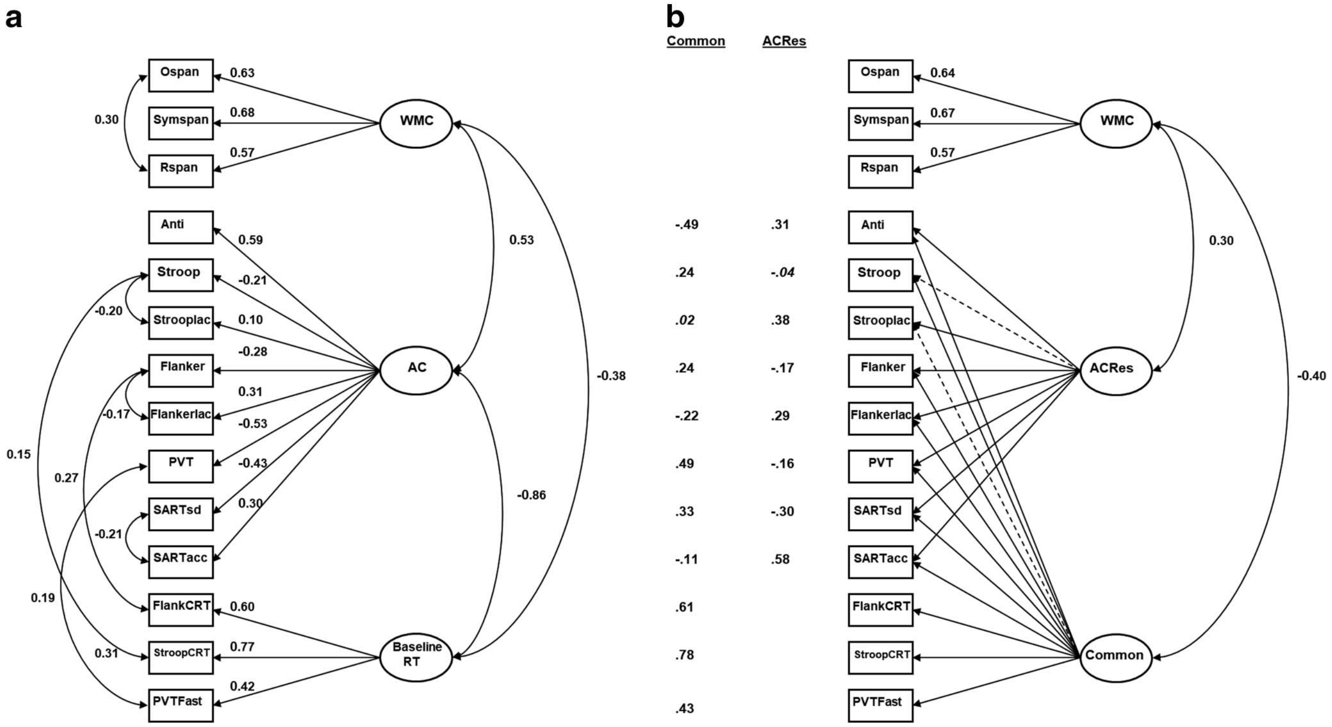
Next, we examined whether the residual variance in AC, after accounting for Baseline RT, is correlated with WMC by examining a bifactor model. In this model, we specified a common factor in which all of the AC and Baseline RT measures loaded onto it. We also specified a residual AC factor in which all of the AC measures loaded onto this factor. These factors were not allowed to correlate with each other but were both allowed to correlate with WMC. Correlations among the residual variances were the same as before. The overall fit of the model was good, $\chi^2(60) = 207.52$, $p < .001$, RMSEA = .028 [.02, .03], CFI = .97, TLI = .95, SRMR = .04. Shown in Figure 4b is the model. Standard errors of the factor loadings were all less than .08. Factor reliability was moderate for all factors (WMC ω = .61, AC residual ω = .41, Baseline RT ω = .60). As can be seen, the common factor was correlated with WMC at −.40 (*SE* = .03). Importantly, the residual AC factor was also correlated with WMC at .30 (*SE* = .05). Thus, even after accounting for shared variance with Baseline RT, most of the AC measures loaded onto the residual AC factor (Stroop did not load significantly), and this residual AC factor was correlated with WMC. Similar results were obtained when using incongruent RTs on the Stroop and flanker tasks instead of the Stroop and flanker RT difference scores, and similar results were obtained when using the Stroop and flanker composite variables. Overall, these results suggest that although some of the relation between WMC and AC is shared with Baseline RT measures, WMC and AC remain correlated even after taking this shared variance into account.

Models Relying Primarily on Reaction Time Difference Scores

The models above relied on various measures of AC, some of which were RT differences scores. As noted previously, and seen in the prior models, these RT difference scores have a number of psychometric issues which can result in lower correlations and less robust results (e.g., Draheim et al., 2019; Hedge et al., 2018; Rouder et al., 2019). Despite these known issues, several studies which have failed to find evidence for a coherent AC factor have primarily relied on difference score measures (e.g., De Simoni & von Bastian, 2018; Gärtner & Strobel, 2019; Rey-Mermet et al., 2018, 2019, 2020). In the next set of analyses, we examined models in which AC was based primarily on RT difference scores from Stroop and flanker along with accuracy on the antisaccade, consistent with prior research. Thus, unlike the prior models we do not include accuracy on Stroop and flanker. In our first model we relied on list-wise deletion for the AC measures to see if they would load on the same factor and be related to a WMC factor. With list-wise deletion on the AC measures there were 714 participants available for this model. The overall fit of the model was good, $\chi^2(7) = 17.10$, $p = .017$, RMSEA = .05 [.02, .07],

Figure 4

(a) Confirmatory Factor Analysis Model for Working Memory Capacity (WMC), Attention Control (AC), and Baseline RT for the Full Combined Dataset (b) Confirmatory Factor Analysis Model for Working Memory Capacity (WMC), Residual Attention Control (ACRes), and a Common Factor for the Full Combined Dataset



Note. The numbers in the Common column represent the factor loadings for each task onto the Common factor; the numbers in ACRes column represent the factor loadings for each task onto the residual AC factor. Paths connecting latent variables (circles) to each other represent the correlations between the constructs, and the numbers from the latent variables to the manifest variables (squares) represent the loadings of each task onto the latent variable. Solid paths are significant at the $p < .05$ level, whereas dashed paths and italicized values are not significant. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; Stroop = Stroop effect; StroopIac = accuracy on incongruent trials in Stroop; Flanker = flanker effect; FlankerIac = accuracy on incongruent trials in Flanker; PVT = psychomotor vigilance task; SARTsd = standard deviation of reaction times (RTs) in sustained attention to response task; SARTacc = accuracy on sustained attention to response task; FlankCRT = RT on congruent trials in Flanker; StroopCRT = RT on congruent trials in Stroop; PVT fast = fastest 20% of RTs on the psychomotor vigilance task.

CFI = .98, TLI = .97, SRMR = .02. Shown in Table 6 is the resulting model. As can be seen, all of the measures loaded significantly on the AC factor (ω = .32), although the loading for Stroop was weak. Furthermore, the AC and WMC factors were correlated (.55). Thus, examining only putative measures of inhibition based primarily on RT difference scores suggested the presence of a coherent AC (or inhibition) factor, and this factor was related to WMC.

Next, to more formally examine the robustness of these results we performed simulations from the overall distribution. Specifically, we took 1,000 random samples of $N = 180$ from the overall distribution of 714 participants. Sample sizes of 180 were chosen based on prior research which has used a similar sample size (e.g., Rey-Mermet et al., 2019). Then, for each down-sampled dataset we specified a two-factor model in which operation span, symmetry span, and reading span loaded onto a WMC factor, and antisaccade, Stroop difference score, and flanker difference score loaded onto an AC factor. These two factors were allowed to correlate. For each model, we saved the resulting parameter estimates and counted the number of instances in which the model did not converge upon a solution. Table 6 shows the results. With this sample size, a model can be expected to converge upon a solution roughly 84% of the time. This may explain why, in some latent variable analyses, attempts to form an AC factor can sometimes fail when relying primarily on difference scores. When the model did converge, the WMC measures tended to load significantly on the model, and both the antisaccade and flanker tended to load on the AC factor most of the time. The Stroop difference score only loaded significantly onto the attention control factor 60% of the time. The correlation between AC and WMC was on average .53, and this relation was significant roughly 93% of the time. We also examined how the results would change with both smaller ($N = 120$) and larger ($N = 360$) sample sizes. As seen in Table 6, when sample sizes were small ($N = 120$) the models tended to converge 76% of the time. The Stroop task loaded significantly on the AC factor only 47% of the time, and flanker loaded significantly only 57% of the time. However, when sample sizes were larger ($N = 360$) the model converged 98% of the time, and all measures loaded significantly on the AC factor over 90% of the time. Thus, sample size seems to have an important influence on whether AC models relying primarily on RT difference scores will converge and whether the measures will load significantly on the AC factor. We contrasted the

Table 6

Average Standardized Factor Loadings, Percentage of Times the Parameter Was Significant, Average Correlations Between Constructs, Percentage of Times the Model Converged, and Average Model Fits for Confirmatory Factor Analyses Simulations

Construct/measure	Full	180 Sim	120 Sim	360 Sim	Acc Sim	Comp Sim
WMC						
Ospan	.84	.78 (100%)	.78 (100%)	.78 (100%)	.78 (100%)	.77 (100%)
Symspan	.57	.61 (100%)	.60 (100%)	.61 (100%)	.61 (100%)	.61 (100%)
Rspan	.81	.74 (100%)	.77 (100%)	.74 (100%)	.75 (100%)	.75 (100%)
AC						
Anti	.49	.50 (92%)	.50 (76%)	.49 (100%)	.57 (96%)	.59 (96%)
Stroop	−.20	−.26 (60%)	−.28 (47%)	−.24 (92%)	.23 (55%)	−.26 (60%)
Flanker	−.36	−.36 (84%)	−.33 (57%)	−.37 (100%)	.45 (95%)	−.45 (96%)
WMC-AC *r*	.52	.55 (93%)	.55 (76%)	.55 (100%)	.41 (88%)	.46 (90%)
% Converge		84%	76%	98%	93%	91%
$\chi^2(7)$		10.81	9.36	12.73	9.66	10.61
CFI		.98	.98	.98	.98	.98
TLI		.97	.98	.99	.98	.98
RMSEA		.04	.04	.03	.03	.03
SRMR		.04	.04	.03	.04	.03

Note. Percentages in parentheses reflect the percentage of times corresponding parameter (i.e., loading) was significant. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade.

RT difference score models with models in which we relied only on accuracy on all three tasks or when we utilized the composite measures for Stroop and flanker (in both Models $N = 180$). As seen in Table 6, these models converged more often than models relying on primarily on RT difference scores. It should be noted that the inclusion of the WMC factor (where all three complex span tasks are intercorrelated and tend to correlate with the AC measures) improves the likelihood of convergence, and the situation may not be as good in studies that do not include a correlated construct or include a less correlated construct (e.g., self-reports of self-control abilities).

We also examined power for the different models with Wang and Rhemtulla's (in press) pwrSEM app. We specified the factor loadings and factor correlation based on the loadings and correlations in the simulated models with 1,000 samples per model. These results suggested that similar to the simulation results, that with smaller sample sizes there was generally insufficient power to reliably detect loadings of the AC measures onto the AC construct. As sample size increased (or accuracy or the composite variables were used) power tended to increase (although power to detect the Stroop loadings were still insufficient) to more sufficient levels. The full results are presented in the Appendix (Table A3). These results suggest that when relying on a fairly large sample size, it is possible to find a coherent AC factor based primarily on RT difference scores from Stroop and flanker along with the antisaccade task. However, this factor is not particularly robust, and relying on much smaller sample sizes of roughly 120–180 participants (typical of many studies, including some of our own) can result less robust results.

Examining Potential Differences Across Sites

One potential issue with mega-analyses of the sort done here is that there might be important differences in the samples, and thus it is not appropriate to simply pool the data together (Curran & Hussong, 2009). In the current analyses we pooled data collected at both the University of Georgia and the University of Oregon. Not only are there geographic differences between these two universities, but there may also be other differences in the samples (such as differences in the ability ranges due to differences in admission criteria). Thus, it is important to examine whether generally similar models are obtained across sites.[6] In particular, it is important to examine measurement invariance across the sites. Measurement invariance refers to whether the same construct is being measured across different groups (Putnick & Bornstein, 2016; Vandenberg & Lance, 2000). In the current study this would be assessing whether AC and WMC are being measured similarly across the different sites. To examine this, we specified a multi-group confirmatory factor analysis with site as our grouping variable. First, we examined configural invariance (i.e., the factor structure is the same across sites) by specifying a model in which

[6] We also examined the multi-level nature of the data further. Because currently multi-level SEM in lavaan is limited to list-wise deletion, it was not possible to examine the full multi-level SEM. Thus, to examine how the results were influenced by Study we first created an AC factor composite for each individual. Next, we constructed a linear mixed-effect model predicting the mean AC factor composite with a random intercept (i.e., a null model). Observations were nested within Study. Using this model, we calculated the intraclass correlation coefficient (ICC), a ratio of between group variance relative to the total variance that ranges from 0.0 to 1.0. Large ICC values indicate a strong relationship between randomly chosen pairs of observations in the same group (i.e., greater degree of dependence attributed to Study). Results revealed a small ICC of .041, meaning 4.1% of the variance in AC scores was between studies. As such, the vast majority of variance in AC scores was at the individual level. However, we should note that small ICC values can give rise to inflated Type I error when one does *not* account for the grouping variable in their analysis—depending on the number of groups and the sample size within each group. To examine the effects of Study on the results, we allowed intercepts to vary by Study. Hence the influence of Study (and the potential Type I error inflation) was effectively zeroed out as far as fixed effects are concerned. Importantly, when adding a WMC composite as a predictor, the linear mixed-effect model specified above revealed a significant moderate relationship between AC and WMC, $\beta = .22$, 95% CI [0.19, 0.26], $t(2524) = 11.69$, $p < .001$. The relation between AC and WMC was the same when not accounting for Study, $\beta = .22$, 95% CI [0.18, 0.26], $t(2525) = 11.31$, $p < .001$. Thus, clustering by Study did not seem to influence the results.

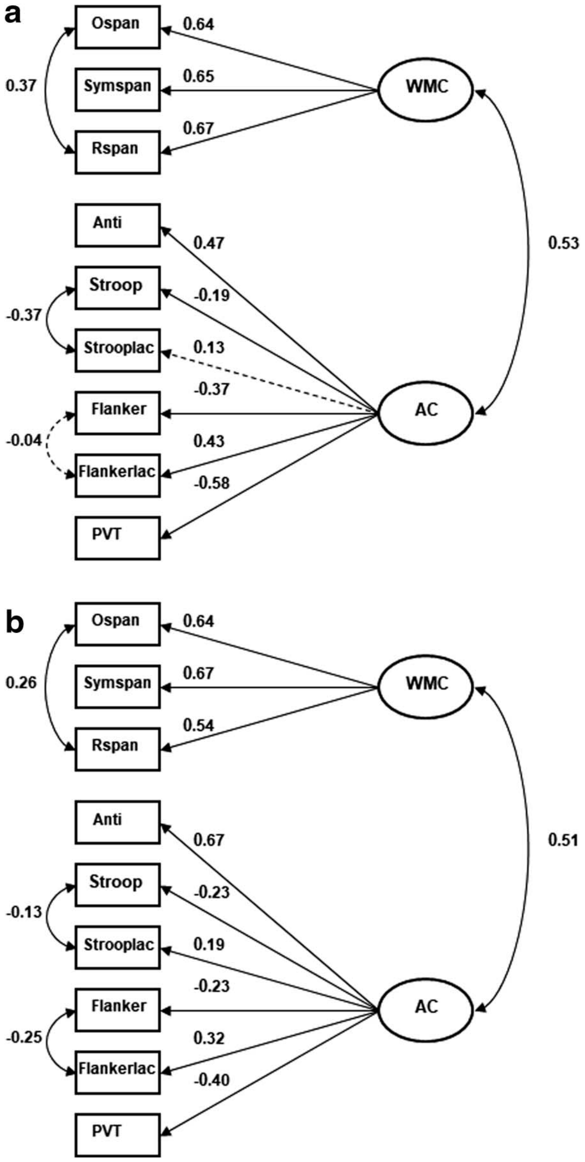
antisaccade, Stroop RT difference score, Stroop incongruent accuracy, flanker RT difference score, flanker incongruent accuracy, and PVT loaded on the AC factor and the three complex span tasks loaded on the WMC factor. Note that we excluded the SART from these models given that it was only administered at the University of Oregon. Residual variances for operation span and reading span were allowed to correlate as were residual variances for Stroop and flanker variables as before. Site was specified as the grouping variable. There were 844 participants from the University of Georgia and 2238 from the University of Oregon. We estimated missing data with full information maximum likelihood. The overall fit of the model was good, $\chi^2(46) = 124.16$, $p < .001$, RMSEA = .03 [.026, .044], CFI = .97, TLI = .95, SRMR = .04, suggesting configural invariance such that the same overall factor structure was seen across sites. Shown in Figure 5 are the models for both the University of Georgia and the University of Oregon separately. As can be seen, many of the factor loadings were similar across sites, and the correlation between AC and WMC was very similar across sites. Next, we examined metric invariance, which is whether the factor loadings across sites are equal. To examine this, we specified the same model as before but now constrained the factor loadings to be equal across sites. The overall fit of the model was good, $\chi^2(53) = 166.35$, $p < .001$, RMSEA = .04 [.031, .044], CFI = .96, TLI = .94, SRMR = .04. To demonstrate metric invariance the fit of the model should not be significantly worse than the configural invariance model. Examining differences in model fit via a difference in χ^2 suggests that the metric invariance model fit worse than the configural invariance model, $\Delta\chi^2(7) = 42.19$, $p < .001$. However, with large samples sizes such as the current study even small differences will tend to be significant. Thus, an examination of the other fit indices is needed. Specifically, CFI, TFI, and RMSEA all dropped by only .01, suggesting that the models were sufficiently similar (Rutkowski & Svetina, 2014). Furthermore, Hildebrandt et al. (2009) suggested that the Root Deterioration per Restriction index (RDR; Browne & Du Toit, 1992) could be used to assess model fit with larger sample sizes with values less than .05 indicating that differences in fit are minor (similar to RMSEA). The RDR index was .04, suggesting that differences in fit were relatively minor. Thus, there was evidence in the data for metric invariance as well. Our final model tested an even more restricted model in which we specified that not only should the factor loadings be equal across sites, but so should the correlation between the factors. The overall fit of the model was good, $\chi^2(54) = 166.46$, $p < .001$, RMSEA = .037 [.03, .043], CFI = .96, TLI = .94, SRMR = .04. As with the prior model, this model fit worse than the configural model using a χ^2 difference test, $\Delta\chi^2(8) = 42.3$, $p < .001$. However, similar to the prior model CFI, TFI, and RMSEA all dropped by only .01 and RDR was .04, suggesting that differences in fit were relatively minor. Thus, overall these results suggest that the model, factor loadings, and factor correlations were generally similar across sites.

Discussion

In the current study data were pooled across multiple prior studies conducted in our laboratory over more than a decade to examine relations among AC measures and whether they are related to WMC. In the large combined dataset we addressed four primary questions. (a) Are AC tasks reliable? (b) Are AC tasks

Figure 5

Confirmatory Factor Analysis Model for Working Memory Capacity (WMC) and Attention Control (AC) for Each Site



Note. (a) Data from University of Georgia. (b) Data from University of Oregon. Paths connecting latent variables (circles) to each other represent the correlations between the constructs and the numbers from the latent variables to the manifest variables (squares) represent the loadings of each task onto the latent variable. Solid paths are significant at the $p < .05$ level, whereas dashed paths are not significant. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; Stroop = Stroop effect; StroopIac = accuracy on incongruent trials in Stroop; Flanker = flanker effect; FlankerIac = accuracy on incongruent trials in Flanker; PVT = psychomotor vigilance task.

related to one another at the zero-order level? (c) Do the AC tasks load onto a general AC factor? And (d) is AC related to WMC? In terms of the first two questions, the results suggested that most of the AC measures had generally acceptable estimates of reliability (although the Stroop RT difference score had poor reliability; average split-half reliability for all AC measures = .74) and most of the AC measures were weakly to moderately correlated with one another. In particular, the average absolute correlation among the AC measures ranged from r = .15 to .20. These results are consistent with two other recent large scale latent variable studies (Redick et al., 2016; average absolute correlation r = .23; Kane et al., 2016; average absolute correlation r = .14) suggesting that various AC measures are weakly to moderately correlated with one another. Given these weak to moderate relations, one might conclude that there is insufficient shared variance across the tasks to suggest a common AC construct. However, we note that similarly weak to moderate relations are seen when comparing various WMC tasks. Specifically, prior research has suggested that span (both complex and simple span) measures of WMC are weakly related to n-back measures of WMC (Redick & Lindsey, 2013). For example, Redick and Lindsey (2013) found that the meta-analytic correlation between complex span and n-back across multiple studies was $r+$ = .20, and the correlations in the individual studies ranged from −.07 to .50. Thus, two putative measures of WMC demonstrate weak to moderate relations. A similar wide range of correlations are seen between complex span measures and visual arrays tasks. For example, Unsworth et al. (2014) reported a correlation of .07 between operation span and visual arrays, whereas Shipstead et al. (2015) reported a correlation of .43 between the same tasks. To get a sense of the overall relation between these two WMC tasks we computed the average correlation across multiple published and unpublished studies which have used these two measures (e.g., Chuderski & Jastrzebski, 2018; Redick et al., 2016; Robison & Unsworth, 2017b; Shipstead et al., 2014, 2015; Unsworth et al., 2014). The resulting average correlation was r = .25 (N = 3099). The meta-analytic correlation was $r+$ = .26. Similar to relations seen between complex span and n-back, measures of complex span and visual arrays tasks demonstrated only moderate relations. These results suggest that disparate measures of WMC tend to be moderately related similar to what is seen with the AC measures. Rather than taking these weak to moderate relations as evidence against general factors, our interpretation is that both AC and WMC represent broad higher-order factors composed of more task/process specific lower-order factors. When factors are represented by a broad collection of tasks then we might expect the relations among the tasks to be weaker than when the factors are composed of largely similar tasks (such as the WMC factor in the current study which is based only on complex span tasks). Overall, we need to be more realistic about the magnitude of correlations (e.g., Funder & Ozer, 2019; Gignac & Szodorai, 2016) across various tasks that rely on a number of processes for performance.

In terms of the third and fourth questions, the results suggested across multiple different models that all of the AC measures loaded significantly on the AC factor and most of the tasks (with the exception of the Stroop variables) loaded moderately well on the AC factor. These loadings remained largely unchanged when various measures were excluded from the models. In particular, when excluding the antisaccade task from the models, the overall loadings remained largely the same suggesting that the AC factor was not simply an antisaccade factor. Excluding other measures resulted in largely similar results. Similarly, relying on list-wise deletion rather than full information maximum likelihood resulted in very similar results. Importantly, in all models, AC and WMC were significantly related around .50 consistent with many prior studies, thereby demonstrating criterion validity for AC. Across many different models the results were remarkably similar in demonstrating consistent factor loading and relations between the AC and WMC factors. Furthermore, WMC was weakly to moderately related to most of the individual AC measures. Collectively, these results provide important evidence for the notion that there is a coherent AC factor which is related to WMC.

Why Are There Discrepancies Across Studies?

Although the current results suggest evidence for AC as a psychometric construct, several recent studies have been unable to find a coherent AC factor (e.g., De Simoni & von Bastian, 2018; Gärtner & Strobel, 2019; Keye et al., 2009; Krumm et al., 2009; Rey-Mermet et al., 2018, 2019, 2020). Naturally, we need to ask why there are discrepancies across studies. That is, why do some studies find evidence for an AC factor, yet other studies find weak or no evidence for an AC factor? As noted previously, one commonality across studies that fail to find an AC factor is that these studies relied predominantly on difference score measures from conflict tasks (either RT difference scores or accuracy difference scores; De Simoni & von Bastian, 2018; Gärtner & Strobel, 2019; Rey-Mermet et al., 2018, 2019, 2020; although see Keye et al., 2009; Rey-Mermet et al., 2018, 2019 for an additional bifactor approach). In each of these studies, more than two thirds of the AC measures were difference scores. In contrast, most of the studies that do find evidence for an AC factor have difference scores as less than a third of measures (an exception to this is Kane et al., 2016 in which many of the measures were difference scores). As noted previously, difference scores tend to have a number of psychometric issues associated with them including the potential for low between-participants variability, poor reliability, and high measurement error which can result in low correlations and an overall inability to find a robust factor (e.g., Draheim et al., 2019; Hedge et al., 2018; Rouder et al., 2019). We examined this notion and found that when relying on small sample sizes and an AC factor composed of only antisaccade accuracy, Stroop RT difference score, and flanker RT difference score that wildly different results can be found with some samples resulting in an AC factor and other samples resulting in near zero correlations among the tasks and an inability to find an AC factor. Indeed, in our simulations we found that roughly 16% of the time the model failed to converge due to low correlations among these AC measures. Furthermore, this AC factor tended to have the lowest estimates of factor reliability. Thus, one potential explanation for discrepancies across studies is that studies that have been unable to find an AC factor primarily relied on difference scores from conflict tasks whereas studies that have found an AC factor relied more on a mix of measures (e.g., Chuderski & Jastrzebski, 2018; Draheim et al., 2019; Friedman et al., 2008; MacKillop et al., 2016; McVay & Kane, 2012; Miyake et al., 2000; Redick et al., 2016; Unsworth & Spillers, 2010; Unsworth & McMillan, 2014; Venables et al., 2018; Von Gunten et al., 2019). As such, low-powered studies that rely primarily on difference scores are unlikely to find a coherent AC factor (see also Rouder et al., 2019). Additionally, by focusing primarily on conflict effects (based on differences scores), it is likely that additional important AC variance (e.g., goal maintenance and fluctuations of attention) is being missed.

There are also likely subtle differences across studies in terms of task construction, samples, instructions to participants, data processing pipelines, and other informal laboratory practices (e.g., Brenninkmeijer et al., 2019). For example, in several studies Rey-Mermet and colleagues (Rey-Mermet et al., 2018, 2019, 2020) have tried to reduce potential influences from episodic memory and associative learning by ensuring that there were no trial-to-trial repetitions of the same stimulus and by counterbalancing aspects of the task including trial types and response keys. Other studies have typically not used these constraints, thus variation in task construction could be a potential reason for discrepancies across studies. Additionally, in several studies Rey-Mermet and colleagues (Rey-Mermet et al., 2018, 2019, 2020) have used a procedure to calibrate target presentation times (i.e., how long the target appeared onscreen before being masked) for each individual's antisaccade trials based on their performance on a prior block of prosaccade trials. Thus, target presentation times for antisaccade trials were different across participants in these studies. In contrast, nearly all other studies have used the same target presentation times for antisaccade trials for all participants. Furthermore, Rey-Mermet and colleagues (Rey-Mermet et al., 2019; Rey-Mermet et al., 2020) have used a difference score between antisaccade and prosaccade trials as the primary antisaccade measure, whereas nearly all other studies have simply used error rates or proportion correct on the antisaccade. Thus, differences in how the tasks are constructed and differences in the measures used could potentially result in discrepancies across studies.

Additionally, some studies that have found a coherent AC factor have tended to use variants of the Stroop task (color-word, spatial, and number Stroop) in which there is a higher proportion of congruent trials than incongruent trials (e.g., the current data; Chuderski & Jastrzebski, 2018; Kane et al., 2016; Redick et al., 2016; Shipstead et al., 2014). Studies that have tended not to find a coherent AC factor, however, have tended to use variants of the Stroop in which proportion congruency is equal (e.g., De Simoni & von Bastian, 2018; Gärtner & Strobel, 2019; Rey-Mermet et al., 2018, 2019, 2020). Previous research has suggested that relations between WMC and the Stroop effect tend to arise in conditions where there is a high proportion of congruent trials relative to incongruent trials, and thus the demands for active goal maintenance are high (Kane & Engle, 2003; Hutchison, 2011; Long & Prat, 2002; Meier & Kane, 2013; Morey et al., 2012). Thus, differences in proportion congruency in the Stroop task could also influence the relations. Although it should be noted that some studies that have found a coherent AC factor used variants of the Stroop with equal numbers of congruent and incongruent trials (e.g., Friedman & Miyake, 2004; Von Gunten et al., 2019). Relatedly, studies that have found weak to near zero correlations between AC measures have also sometimes demonstrated smaller Stroop and flanker effects than studies that tend to find correlations between AC measures. For example, Hedge et al. (2018) reported Stroop effects ranging from 61–91 *ms* (*SD*s 34–51) and flanker effects ranging from 35–46 *ms* (*SD*s 34–51). Similarly, Gärtner and Strobel (2019) reported Stroop and flanker effects around 47 *ms* (*SD*s 39–52). However, as seen in Table 1 in the current study, the effects were much larger (Stroop = 148 *ms, SD* = 97; flanker = 114 *ms, SD* = 68). Large effects are seen in other studies demonstrating correlations among the AC measures (e.g., Friedman & Miyake, 2004; Redick et al., 2016; Shipstead et al., 2014; Von Gunten et al., 2019). Again, differences in task construction (such as total number of trials and proportion congruency) could be influencing the magnitude of the effects as well as the amount of between-subjects variability which could then influence relations among the AC tasks.

As mentioned above, another potential difference that could give rise to discrepancies across studies are the samples used. In particular, some studies that have failed to find a coherent AC factor have utilized students from relatively selective universities (e.g., De Simoni & von Bastian, 2018; Rey-Mermet et al., 2019; Rey-Mermet et al., 2020), whereas some studies that have found a coherent AC factor have utilized students from less selective universities (e.g., the current data; Chuderski & Jastrzebski, 2018; Kane et al., 2016; Von Gunten et al., 2019) or have used a combination of students and community participants (e.g., Draheim et al., 2020; Redick et al., 2016; Shipstead et al., 2014). When participants are sampled from a restricted range, abilities tend to make a small contribution to any observed correlations (e.g., Deary et al., 1996). Instead, the variability that is detected is more likely due to task-specific skills. Thus, the ability to find already small relations is reduced even further. Although it should be noted that some studies that have failed to find a coherent AC factor do find typical strong latent correlations between WMC and fluid intelligence (e.g., De Simoni & von Bastian, 2018; Rey-Mermet et al., 2019), suggesting there is enough variability in abilities to detect these relations. Future research is needed to examine how potential differences in sample characteristics influence the ability to find a coherent AC factor.

There seem to be a number of differences between studies that have consistently demonstrated a coherent AC factor versus those that have failed to find a coherent AC factor. It is unlikely that any one factor is responsible for discrepancies across studies. As such, it may be difficult to pinpoint potential reasons for discrepant results as multiple factors could be at play. What seems needed is an adversarial collaboration between groups that typically find an AC factor and groups that typically fail to find such a factor. In such a study, participants from different sites would perform a large battery of AC, WMC, and potentially other tasks. Different variants of AC tasks would be given to examine how task construction and different measures influence the ability to find a coherent AC factor and whether differences in samples influence the relations.

Despite discrepancies across some studies, the overall bulk of evidence across many latent variable studies suggests the presence of a coherent AC factor that is related to WMC. Future research should be mindful that factor analytic studies of AC are still in their infancy compared with other cognitive abilities. As such, additional research is needed to better understand the structure of AC abilities.

Limitations and Future Directions

The current results provide important information on the nature of AC as an individual differences construct. At the same time, there are a number of limitations which need to be addressed. For example, all of the current data come from studies conducted in our laboratory over the last 10 plus years, and thus this is not a comprehensive analysis of all existing studies that have been done on this topic (as you might find in a meta-analysis). There are clear benefits in doing mega-analyses of the type done in the current study. These include increased sample sizes, greater power, and greater precision in detecting the effects of interest. At the same time there are limitations in mega-analyses in that pooling data across multiple different studies can lead to heterogeneity across the samples and measures (Curran & Hussong, 2009). For example, although similar samples of participants were used in each study and each participant performed roughly the same set of AC and WMC tasks, there are still likely differences in the studies in terms of when participants were tested and the exact measures they were tested on. That is, in nearly all of the studies participants completed other measures including measures of long-term memory and fluid intelligence. Additionally, task orders were not fully consistent across studies (although in most of the studies the tasks were given in the same 2-hr session). It is not known how these factors could impact the results, and thus it is possible that even with large sample sizes, there is some noise in the relations attributable to this heterogeneity. We attempted to examine this issue by examining differences in the factor structure across sites (University of Georgia vs. University or Oregon) and found that for the most part there was evidence for measurement invariance across the sites, suggesting that AC and WMC were measured similarly at both universities. Of course, there could still be important differences that could influence the results. Thus, the current mega-analyses are just one means of synthesizing a large amount of data. Future latent variable studies with large sample sizes and a large number of measures are needed to replicate and extend the current results.

An additional limitation of the current study is that nearly all of the participants were college students at two comprehensive state universities. Thus, the current samples are likely composed of more high ability participants than is ideal. As noted previously, this partial range restriction of the ability distributions likely resulted in weaker correlations than would be seen with a broader ability range. For example, Redick et al. (2016) consisted of data from several universities as well as community volunteers from the Atlanta metro area. Correlations in this study tended to be larger than those seen in the current data, suggesting that larger relations will likely be seen when utilizing a broader ability range (see also Draheim et al., 2020). When examining relations among AC tasks and their relation with WMC, future research should include a broader range of abilities when possible.

Another limitation of the current study is that we did not fully assess how processing speed or speed–accuracy trade-offs could influence the measurement of AC. Specifically, some prior research has suggested that AC measures are confounded with processing speed (e.g., Jewsbury et al., 2016; Rey-Mermet et al., 2019) and are influenced by speed–accuracy trade-offs (Draheim et al., 2019, 2020; Rey-Mermet et al., 2019) which can impact the measurement of AC. We examined the potential role of processing speed by including a Baseline RT factor composed of congruent RTs on the Stroop and flanker tasks along with the fastest RTs on the psychomotor vigilance task. This latent factor was strongly related to AC, but demonstrated weaker relations with WMC. In particular, AC was more strongly related to WMC than the Baseline RT factor was. We further examined these relations via a bifactor model in which all of the AC measures and the Baseline RT measured loaded onto a common factor, and the AC measures also loaded onto a residual AC factor. The results suggested that both factors were correlated with WMC. That is, variance in AC was significantly related to WMC, even when taking into account shared variance with Baseline RT. Thus, these results suggest that some of the relation between WMC and AC might be attributable to shared variance with something like processing speed, but there is considerable shared variance between WMC and AC when accounting for Baseline RT. However, we note that the Baseline RT factor is not necessarily a pure measure of processing speed as other factors such as fluctuations of attention and goal neglect can influence these Baseline RT measures (e.g., Kane et al., 2016; Unsworth, 2015). Thus, some of the shared variance between WMC and the Baseline RT factor could be due to factors other than just processing speed. Future research is needed to better examine these relations with independent processing speed measures.

It is also important to note that the specific goal in the current study was to examine relations among AC measures and their relations with WMC as they have been typically measured in the past. We are in general agreement with other researchers in suggesting that other constructs such as processing speed need to be examined and that strategic choices (such as speed-accuracy decisions) during AC tasks (and all other tasks) need to be examined as additional important sources of variability. At the same time, we note that no task (and probably no factor) is process pure. That is, a number of processes are likely influencing performance on AC and WMC tasks (as well as fluid intelligence tasks). These include AC, working memory, processing speed, long-term memory, self-efficacy, task engagement, speed–accuracy trade-off decisions, task-specific motivation, as well as task-specific strategies to name a few. The extent to which a collection of similar processes are influencing performance across similar tasks will result in a factor that is similarly the result of multiple processes (e.g., Detterman, 1994; Kovacs & Conway, 2016). These issues are not isolated only to AC tasks, as prior research has suggested that processing speed and speed–accuracy trade-offs are likely important for performance on WMC and fluid intelligence tasks (e.g., Ackerman & Ellingsen, 2016; Chuderski, 2013; Daneman & Tardif, 1987; Kyllonen, 1994; Kyllonen & Zu, 2016; Lohman, 1989; Phillips & Rabbitt, 1995; Salthouse, 1996; Wilhelm & Schulze, 2002). Furthermore, prior research has suggested that processing speed measures (like many measures) are influenced by AC abilities (Carroll, 1993; Cepeda et al., 2013; Horn & Blankson, 2005; Lustig et al., 2006; Schneider & McGrew, 2012, 2018). Future research is needed to examine how various processes contribute to performance on AC and WMC tasks and how these processes influence the relation between the two constructs. Furthermore, we are in agreement with recent proposals suggesting the need to develop more reliable and valid measures of AC.

A final limitation is that in the Introduction we suggested that broad AC abilities could be broken down into more specific abilities such as restraining, constraining, and sustaining attention and measures of each more specific ability were used in our broad AC factor. However, we did not explicitly examine whether such a breakdown of the AC factor is possible. That is, we did not test whether there are three (and potentially more) AC subfactors which are correlated with one another and load onto a higher-order AC factor. We could not test such a model with the current data given that only one task thought to measure constraining attention (flanker) was used. To fully test such a notion, multiple measures of each specific ability would be needed. Furthermore, we note that some prior evidence suggests the possibility of these distinct factors. As noted in the Introduction, Kane et al. (2016) found evidence for distinct restraint and constraint factors that were both related to WMC (reanalyses of Redick et al., 2016 suggested similar results). Additionally, in a recent study from our lab (Unsworth et al., 2020; these data are included in the current mega-analysis) we found evidence for a sustained attention factor (based on lapses of attention) that was related to a restraint/constraint factor (−.69) and to WMC (−.34). Thus, there is some evidence for distinct, yet related AC subfactors. However, no study to date has examined all three factors at once. Future research should better examine the notion that there are distinct AC abilities and how these abilities are related to other constructs.

Conclusions

Collectively the current results suggest that AC measures are weakly to moderately correlated with one another and all load onto the same general AC factor. This factor was correlated with WMC consistent with much prior research and theorizing. The current mega-analyses provide important evidence on the nature of a general AC ability, which is related to other critical cognitive abilities. Future research is needed to further delineate the nature of AC abilities and place them into the broader context of cognitive abilities.

References

Ackerman, P. L., & Ellingsen, V. J. (2016). Speed and accuracy indicators of test performance under different instructional conditions: Intelligence correlates. *Intelligence*, *56*, 1–9. https://doi.org/10.1016/j.intell.2016.02.004

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*, 525–531. https://doi.org/10.1017/S1366728909990423

Blettner, M., Sauerbrei, W., Schlehofer, B., Scheuchenpflug, T., & Friedenreich, C. (1999). Traditional reviews, meta-analyses and pooled analyses in epidemiology. *International Journal of Epidemiology*, *28*, 1–9. https://doi.org/10.1093/ije/28.1.1

Brenninkmeijer, J., Derksen, M., & Rietzschel, E. (2019). Informal laboratory practices in psychology. *Collabra Psychology*, *5*, 45. https://doi.org/10.1525/collabra.221

Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language*, *66*, 407–415. https://doi.org/10.1016/j.jml.2011.12.009

Browne, M. W., & Du Toit, S. H. (1992). Automated fitting of nonstandard models. *Multivariate Behavioral Research*, *27*, 269–300. https://doi.org/10.1207/s15327906mbr2702_13

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Theory and methodology of frontal and executive function* (pp. 81–116). Psychology Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Cepeda, N. J., Blackwell, K. A., & Munakata, Y. (2013). Speed isn't everything: Complex processing speed measures mask individual differences and developmental changes in executive control. *Developmental Science*, *16*, 269–286. https://doi.org/10.1111/desc.12024

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence*, *41*, 244–262. https://doi.org/10.1016/j.intell.2013.04.003

Chuderski, A., & Jastrzębski, J. (2018). Much ado about aha! Insight problem solving is strongly related to memory capacity and reasoning ability. *Journal of Experimental Psychology: General*, *147*, 257–281. https://doi.org/10.1037/xge0000378

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*, 278–295. https://doi.org/10.1016/j.intell.2012.02.010

Colflesh, G. J. H., & Conway, A. R. A. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review*, *14*, 699–703. https://doi.org/10.3758/BF03196824

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*, 331–335. https://doi.org/10.3758/BF03196169

Costafreda, S. G. (2009). Pooling fMRI data: Meta-analysis, mega-analysis and multi-center studies. *Frontiers in Neuroinformatics*, *3*, 1–8. https://doi.org/10.3389/neuro.11.033.2009

Curran, P. J., Cole, V., Giordano, M., Georgeson, A. R., Hussong, A. M., & Bauer, D. J. (2018). Advancing the study of adolescent substance use through the use of integrative data analysis. *Evaluation & the Health Professions*, *41*, 216–245. https://doi.org/10.1177/0163278717747947

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*, 81–100. https://doi.org/10.1037/a0015914

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29. https://doi.org/10.1037/1082-989X.1.1.16

Daneman, M., & Tardif, T. (1987). Working memory and reading skill reexamined. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 491–508). Erlbaum.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*, 105–132. https://doi.org/10.1016/S0160-2896(96)90008-2

De Simoni, C., & von Bastian, C. C. (2018). Working memory updating and binding training: Bayesian evidence supporting the absence of transfer. *Journal of Experimental Psychology: General*, *147*, 829–858. https://doi.org/10.1037/xge0000453

Detterman, D. K. (Ed.). (1994). Theoretical possibilities: The relation of human intelligence to basic cognitive abilities. *Current topics in human intelligence, Vol. 4: Theories of intelligence* (pp. 85–115). Ablex.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers*, *17*, 652–655. https://doi.org/10.3758/BF03200977

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. https://doi.org/10.1037/bul0000192

Draheim, C., Tsukahara, J. S., Martin, J., Mashburn, C., & Engle, R. W. (2020). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*. Advance online publication. https://doi.org/10.1037/xge0000783

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, *30*, 257–303. https://doi.org/10.1006/cogp.1996.0008

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two- factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). Elsevier.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*, 340–347. https://doi.org/10.1162/089892902317361886

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. https://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*, 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*, 156–168. https://doi.org/10.1177/2515245919847202

Gärtner, A., & Strobel, A. (2019). *Individual differences in inhibitory control: A latent variable analysis*. Retrieved from https://doi.org/10.31234/osf.io/gnhmt

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). Academic Press.

Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predicts susceptibility to interference in verbal working memory. *Neuropsychology*, *20*, 511–528. https://doi.org/10.1037/0894-4105.20.5.511

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. https://doi.org/10.3758/s13428-017-0935-1

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, *136*, 217–240. https://doi.org/10.1037/0096-3445.136.2.217

Hildebrandt, A., Wilhelm, O., & Robitzsch, A. (2009). Complementary andcompeting factor analytic approaches for the investigation of measurement invariance. *Review of Psychology*, *16*, 87–102.

Himi, S. A., Buhner, M., Schwaighofer, M., Klapetek, A., & Hilbert, S. (2019). Multitasking behavior and its related constructs: Executive functions, working memory capacity, relational integration, and divided attention. *Cognition*, *189*, 275–298. https://doi.org/10.1016/j.cognition.2019.04.010

Horn, J. L., & Blankson, N. (2005). Foundations for better under-standing of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (2nd ed., pp. 41–68). Guilford Press.

Hussong, A. M., Flora, D. B., Curran, P. J., Chassin, L. A., & Zucker, R. A. (2008). Defining risk heterogeneity for internalizing symptoms among children of alcoholic parents: A prospective cross-study analysis. *Development and Psychopathology*, *20*, 165–193. https://doi.org/10.1017/S0954579408000084

Hutchison, K. A. (2011). The interactive effects of listwide control, item based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 851–860. https://doi.org/10.1037/a0023437

James, A. N., Fraundorf, S. H., Lee, E-K., & Watson, D. G. (2018). Individual differences in syntactic processing: Is there evidence for reader-text interactions? *Journal of Memory and Language*, *102*, 155–181. https://doi.org/10.1016/j.jml.2018.05.006

Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell—Horn—Carroll model. *Journal of Experimental Psychology: General*, *145*, 220–245. https://doi.org/10.1037/xge0000119

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

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. https://doi.org/10.3758/BF03196323

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. https://doi.org/10.1037/0096-3445.132.1.47

Kane, M. J., Gross, G. M., Chun, C. A., Smeekens, B. A., Meier, M. E., Silvia, P. J., & Kwapil, T. R. (2017). For whom the mind wanders, and when, varies across laboratory and daily-life settings. *Psychological Science*, *28*, 1271–1289. https://doi.org/10.1177/0956797617706086

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General*, *145*, 1017–1048. https://doi.org/10.1037/xge0000184

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*, 1147–1185. https://doi.org/10.1037/bul0000160

Keye, D., Wilhelm, O., Oberauer, K., & Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: Testing means and covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychological Research*, *73*, 762–776. https://doi.org/10.1007/s00426-008-0188-9

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*, 151–177. https://doi.org/10.1080/1047840X.2016.1153946

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence*, *37*, 347–364. https://doi.org/10.1016/j.intell.2009.02.003

Kyllonen, P. C. (1994). CAM: A theoretical framework for cognitive abilities measurement. In D. K. Detterman (Ed.), *Current topics in human intelligence, Vol. 4: Theories of intelligence* (pp. 307–359). Ablex.

Kyllonen, P. C., & Zu, J. (2016). Use of response time for measuring cognitive ability. *Journal of Intelligence*, *4*, 14. https://doi.org/10.3390/jintelligence4040014

Liesefeld, H. R., Fu, X., & Zimmer, H. D. (2015). Fast and careless or careful and slow? Apparent holistic processing in mental rotation is explained by speed–accuracy trade-offs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1140–1151. https://doi.org/10.1037/xlm0000081

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy trade-offs. *Behavior Research Methods*, *51*, 40–60. https://doi.org/10.3758/s13428-018-1076-x

Lohman, D. F. (1989). Estimating individual differences in information processing using speed-accuracy models. In Kanfer, R., Ackerman, P. L., Cudeck, R., (Eds.), *Abilities, motivation, methodology: The Minnesota Symposium on Learning and Individual Differences* (pp. 119–163). Erlbaum.

Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition*, *30*, 294–301. https://doi.org/10.3758/BF03195290

Lustig, C., Hasher, L., & Tonev, S. T. (2006). Distraction as a determinant of processing speed. *Psychonomic Bulletin & Review*, *13*, 619–625. https://doi.org/10.3758/BF03193972

MacKillop, J., Weafer, J., Gray, J. C., Oshri, A., Palmer, A., & de Wit, H. (2016). The latent structure of impulsivity: Impulsive choice, impulsive action, and impulsive personality traits. *Psychopharmacology*, *233*, 3361–3370. https://doi.org/10.1007/s00213-016-4372-0

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 196–204. https://doi.org/10.1037/a0014104

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. https://doi.org/10.1037/a0025250

Meier, M. E., & Kane, M. J. (2013). Working memory capacity and Stroop interference: Global versus local indices of executive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 748–759. https://doi.org/10.1037/a0029200

Meier, M. E., & Kane, M. J. (2015). Carving executive control at its joints: Working memory capacity predicts stimulus-stimulus, but not stimulus-response, conflict. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1849–1872. https://doi.org/10.1037/xlm0000147

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. https://doi.org/10.1006/cogp.1999.0734

Morey, C. C., Elliott, E. M., Wiggers, J., Eaves, S. D., Shelton, J. T., & Mall, J. T. (2012). Goal-neglect links Stroop interference with working memory capacity. *Acta Psychologica*, *141*, 250–260. https://doi.org/10.1016/j.actpsy.2012.05.013

Paap, K. R., Anders-Jefferson, R., Zimiga, B., Mason, L., & Mikulinsky, R. (2020). Interference scores have inadequate concurrent and convergent validity: Should we stop using the flanker, Simon, and spatial Stroop tasks? *Cognitive Research: Principles and Implications*, *5*, 7. https://doi.org/10.1186/s41235-020-0207-y

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. https://doi.org/10.1016/j.jneumeth.2016.10.002

Pettigrew, C., & Martin, R. C. (2014). Cognitive declines in healthy aging: Evidence from multiple aspects of interference resolution. *Psychology and Aging*, *29*, 187–204. https://doi.org/10.1037/a0036085

Phillips, L. H., & Rabbitt, P. M. A. (1995). Impulsivity and speed-accuracy strategies in intelligence test performance. *Intelligence*, *21*, 13–29. https://doi.org/10.1016/0160-2896(95)90036-5

Poole, B. J., & Kane, M. J. (2009). Working memory capacity predicts the executive control of visual search among distractors: The influence of sustained and selective attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *62*, 1430–1454. https://doi.org/10.1080/17470210802479329

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Rabbitt, P. (Ed.). (1997). Introduction: Methodologies and models in the study of executive function. *Methodology of frontal and executive function* (pp. 1–38). Psychology Press.

Raykov, T. (2001a). Bias of coefficient alpha for fixed congeneric measures with correlated errors. *Applied Psychological Measurement*, *25*, 69–76. https://doi.org/10.1177/01466216010251005

Raykov, T. (2001b). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, *54*, 315–323. https://doi.org/10.1348/000711001159582

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Redick, T. S. (2014). Cognitive control in context: Working memory capacity and proactive control. *Acta Psychologica*, *145*, 1–9. https://doi.org/10.1016/j.actpsy.2013.10.010

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: Selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 308–324. https://doi.org/10.1037/a0022216

Redick, T. S., & Engle, R. W. (2006). Working memory capacity and attention network test performance. *Applied Cognitive Psychology*, *20*, 713–721. https://doi.org/10.1002/acp.1224

Redick, T. S., & Engle, R. W. (2011). Integrating working memory capacity and context-processing views of cognitive control. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *64*, 1048–1055. https://doi.org/10.1080/17470218.2011.577226

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *20*, 1102–1113. https://doi.org/10.3758/s13423-013-0453-9

Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., Kane, M. J., Hambrick, K. L., & Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General*, *145*, 1473–1492. https://doi.org/10.1037/xge0000219

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 501–526. https://doi.org/10.1037/xlm0000450

Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General*, *148*, 1335–1372. https://doi.org/10.1037/xge0000593

Rey-Mermet, A., Singh, K., Gignac, G. E., Brydges, C., & Ecker, U. K. H. (2020). *Removal of information from working memory is not related to inhibition*. https://doi.org/10.31234/osf.io/hdks9

Richmond, L., Redick, T. S., & Braver, T. (2016). Remembering to prepare: The benefits (and costs) associated with high working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1764–1777.

Robison, M. K., Miller, A. L., & Unsworth, N. (in press). A multifaceted approach to understanding individual differences in mind-wandering. *Cognition*.

Robison, M. K., & Unsworth, N. (2016). Do participants differ in their cognitive abilities, task motivation, or personality characteristics as a function of time of participation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 897–913. https://doi.org/10.1037/xlm0000215

Robison, M. K., & Unsworth, N. (2017a). Individual differences in working memory capacity predict learned control over attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1912–1924. https://doi.org/10.1037/xhp0000419

Robison, M. K., & Unsworth, N. (2017b). Variation in the use of cues to guide attention in visual working memory. *Attention, Perception, & Psychophysics*, *79*, 1652–1665. https://doi.org/10.3758/s13414-017-1335-4

Robison, M. K., & Unsworth, N. (2018). Cognitive and contextual correlates of spontaneous and deliberate mind-wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 85–98. https://doi.org/10.1037/xlm0000444

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

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

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, *74*, 31–57. https://doi.org/10.1177/0013164413498257

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428. https://doi.org/10.1037/0033-295X.103.3.403

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*, 532–545. https://doi.org/10.1037/0894-4105.19.4.532

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*, 566–594. https://doi.org/10.1037/0096-3445.132.4.566

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage.

Savalei, V., & Reise, S. P. (2019). Don't forget the model in your model-based reliability coefficients: A reply to McNeish. *Collabra: Psychology*, *5*, 36.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*, 23–74.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carrol model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). Guilford Press.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carrol model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). Guilford Press.

Schubert, A-L., & Rey-Mermet, A. (2019). Does Process Overlap Theory replace the issues of general intelligence with the issues of attentional control? *Journal of Applied Research in Memory & Cognition*, *8*, 277–283. https://doi.org/10.1016/j.jarmac.2019.06.004

Scoboria, A., Wade, K. A., Lindsay, D. S., Azad, T., Strange, D., Ost, J., & Hyman, I. E. (2017). A mega-analysis of memory reports from eight peer-reviewed false memory implantation studies. *Memory*, *25*, 146–163. https://doi.org/10.1080/09658211.2016.1260747

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics*, *77*, 1863–1880. https://doi.org/10.3758/s13414-015-0899-0

Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116–141. https://doi.org/10.1016/j.jml.2014.01.004

Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143*, 850–886. https://doi.org/10.1037/a0033981

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence*, *49*, 110–128. https://doi.org/10.1016/j.intell.2015.01.005

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin*, *145*, 79–139. https://doi.org/10.1037/bul0000176

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2012). Variation in cognitive failures: An individual differences investigation of everyday attention and memory failures. *Journal of Memory and Language*, *67*, 1–16. https://doi.org/10.1016/j.jml.2011.12.005

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

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

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., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica*, *150*, 14–25. https://doi.org/10.1016/j.actpsy.2014.04.001

Unsworth, N., & McMillan, B. D. (2017). Attentional disengagements in educational contexts: A diary investigation of everyday mind-wandering and distraction. *Cognitive Research: Principles and Implications*, *2*, 32. https://doi.org/10.1186/s41235-017-0070-7

Unsworth, N., McMillan, B. D., Brewer, G. A., & Spillers, G. J. (2012). Everyday attention failures: An individual differences investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1765–1772. https://doi.org/10.1037/a0028075

Unsworth, N., Miller, A. L., & Robison, M. K. (2020). Individual differences in lapses of sustained attention: Oculometric indicators of intrinsic alertness. *Journal of Experimental Psychology: Human Perception and Performance*, *46*, 569–592. https://doi.org/10.1037/xhp0000734

Unsworth, N., Miller, J. D., Lakey, C. E., Young, D. L., Meeks, J. T., Campbell, W. K., & Goodie, A. S. (2009). Exploring the relations among executive functions, fluid intelligence, and personality. *Journal of Individual Differences*, *30*, 194–200. https://doi.org/10.1027/1614-0001.30.4.194

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J., & 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. https://doi.org/10.1080/09658210902998047

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, *38*, 111–122. https://doi.org/10.1016/j.intell.2009.08.002

Unsworth, N., Redick, T. R., McMillan, B. D., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2015). Is playing videogames related to cognitive abilities? *Psychological Science*, *26*, 759–774. https://doi.org/10.1177/0956797615570367

Unsworth, N., Redick, T. S., Spillers, G. J., & Brewer, G. A. (2012). Variation in working memory capacity and cognitive control: Goal maintenance and micro-adjustments of control. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *65*, 326–355. https://doi.org/10.1080/17470218.2011.597865

Unsworth, N., & Robison, M. K. (2017a). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1962–1987. https://doi.org/10.1037/xlm0000421

Unsworth, N., & Robison, M. K. (2017b). A Locus Coeruleus-Norepinephrine account of individual differences in working memory capacity and attention control. *Psychonomic Bulletin & Review*, *24*, 1282–1311. https://doi.org/10.3758/s13423-016-1220-5

Unsworth, N., & Robison, M. K. (2020). Working memory capacity and sustained attention: A cognitive-energetic perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*, 77–103. https://doi.org/10.1037/xlm0000712

Unsworth, N., Robison, M. K., & Miller, A. L. (2019). Individual differences in baseline oculometrics: Examining variation in baseline pupil diameter, spontaneous eye blink rate, and fixation stability. *Cognitive, Affective & Behavioral Neuroscience*, *19*, 1074–1093. https://doi.org/10.3758/s13415-019-00709-z

Unsworth, N., Robison, M. K., & Miller, A. L. (2020). *Individual differences in lapses of attention: A latent variable analysis*. Manuscript submitted for publication.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1302–1321. https://doi.org/10.1037/0278-7393.30.6.1302

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention, Memory, or Both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*, 392–406. https://doi.org/10.1016/j.jml.2010.02.001

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2009). Examining the relations among working memory capacity, attention control, and fluid intelligence from a dual-component framework. *Psychology Science*, *51*, 388–402.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70. https://doi.org/10.1177/109442810031002

van Zomeren, A. H., & Brouwer, W. H. (1994). *Clinical neurospychology of attention*. Oxford Press.

Venables, N. C., Foell, J., Yancey, J. R., Kane, M. J., Engle, R. W., & Patrick, C. J. (2018). Quantifying inhibitory control as externalizing proneness: A cross-domain model. *Clinical Psychological Science*, *6*, 561–580. https://doi.org/10.1177/2167702618757690

Von Gunten, C. D., Bartholow, B. D., & Martins, J. S. (2019). *Inhibition tasks are not associated with self-regulation outcomes in healthy college students*. https://doi.org/10.31234/osf.io/uwbdg

Wang, Y. A., & Rhemtulla, M. (in press).Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science*.

Was, C. A. (2007). Further evidence that not all executive functions are equal. *Advances in Cognitive Psychology*, *3*, 399–407. https://doi.org/10.2478/v10053-008-0004-5

Weldon, R. B., Mushlin, H., Kim, B., & Sohn, M. H. (2013). The effect of working memory capacity on conflict monitoring. *Acta Psychologica*, *142*, 6–14. https://doi.org/10.1016/j.actpsy.2012.10.002

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433. https://doi.org/10.3389/fpsyg.2013.00433

Wilhelm, O., & Schulze, R. (2002). The relation of speeded and unspeeded reasoning with mental speed. *Intelligence*, *30*, 537–554. https://doi.org/10.1016/S0160-2896(02)00086-7

(*Appendix follows*)

Appendix

Additional Models and Estimates of Power

Models With Composite Variables and List-Wise Deletion

We examined the same series of list-wise deletion models as seen in Table 5 to see whether similar overall results are found when utilizing the composite Stroop and flanker variables. First, we examined relations among antisaccade, flanker, and PVT, and we specified these three measures to load on the AC factor (labeled AFP), while the three working memory measures loaded on the WMC factor. With list-wise deletion there were 1018 participants available for this model. The overall fit of the model was good, $\chi^2(7) = 8.48$, $p = .29$, RMSEA = .01 [.00, .04], CFI = .99, TLI = .99, SRMR = .01. Next, we specified a model in which antisaccade, the Stroop composite, and the flanker composite all loaded onto the AC (or inhibition) factor (labeled ASF). The overall fit of the model was good, $\chi^2(7) = 13.98$, $p = .052$, RMSEA = .04 [.00, .066], CFI = .99, TLI = .98, SRMR = .02. Finally, we examined a model in which antisaccade, Stroop, and PVT loaded on the AC factor (labeled ASP). The overall fit of the model was good, $\chi^2(7) = 20.38$, $p = .005$, RMSEA = .04 [.02, .06], CFI = .99, TLI = .97, SRMR = .02. Shown in Table A1 are the results from the models. In each model, all of the AC measures loaded significantly on the AC factor, and AC and WMC were correlated. Collectively, these results suggest that using composite variables that combine RT difference scores and incongruent accuracy on Stroop and flanker resulted in higher factor loadings for these measures on the AC factor, and the overall AC factor was consistently related to WMC.

Table A1

Standardized Factor Loadings, Standard Errors, and Correlations Between Constructs for Confirmatory Factor Analyses With Stroop and Flanker Composite Variables and List-Wise Deletion

Construct/measure	AFP	ASF	ASP
WMC			
Ospan	.68 (.05)	.83 (.07)	.64 (.04)
Symspan	.67 (.05)	.58 (.06)	.71 (.04)
Rspan	.63 (.05)	.80 (.08)	.48 (.04)
AC			
Antisaccade	.52 (.04)	.56 (.06)	.60 (.04)
StroopC		−.24 (.05)	−.26 (.04)
FlankerC	−.44 (.04)	−.47 (.05)	
PVT	−.50 (.04)		−.41 (.04)
WMC-AC *r*	.58 (.05)	.46 (.07)	.55 (.05)

Note. Standard errors are in parentheses. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; PVT = psychomotor vigilance task; StroopComp = composite variable combining RT Stroop effect with incongruent accuracy; FlankerComp = composite variable combining RT Flanker effect with incongruent accuracy; AFP = antisaccade, flanker, and psychomotor vigilance task model; ASF = antisaccade, Stroop, and flanker model; ASP = antisaccade, Stroop, and psychomotor vigilance task model.

(*Appendix continues*)

Table A2

Standardized Factor Loadings, Standard Errors, Correlations Between Constructs, and Model Fits for Confirmatory Factor Analyses for Alternative Models

Construct/ measure	MLR	Transformed	Outlier
WMC			
Ospan	.63 (.03)	.63 (.03)	.62 (.03)
Symspan	.68 (.03)	.68 (.03)	.69 (.03)
Rspan	.56 (.03)	.56 (.03)	.55 (.03)
AC			
Antisaccade	.59 (.03)	.63 (.02)	.64 (.03)
Stroop	−.21 (.04)	−.20 (.03)	−.17 (.03)
StroopIAcc	.18 (.04)	.15 (.03)	.16 (.03)
Flanker	−.28 (.04)	−.26 (.03)	−.24 (.03)
FlankerIacc	.34 (.04)	.37 (.03)	.29 (.03)
PVT	−.47 (.03)	−.57 (.02)	−.50 (.03)
SARTsd	−.46 (.05)	−.46 (.04)	−.42 (.04)
SARTAcc	.40 (.04)	.40 (.04)	.40 (.04)
WMC-AC *r*	.53 (.03)	.50 (.03)	.49 (.03)
χ^2 (39)	106.28	150.21	146.50
CFI	.97	.96	.96
TLI	.96	.95	.95
RMSEA	.03	.03	.03
SRMR	.04	.04	.04

Note. Standard errors are in parentheses. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade; Stroop = RT Stroop effect; StroopIAcc = accuracy on incongruent trials in Stroop; Flanker = RT flanker effect; FlankerIAcc = accuracy on incongruent trials in Flanker; PVT = psychomotor vigilance task; SARTsd = standard deviation of reaction times in sustained attention to response task; SARTacc = accuracy on sustained attention to response task.

Alternative Confirmatory Factor Analyses

Given potential issues with skewed measures and potential outliers, here we present several alternative confirmatory factor analyses. As will be seen, all of the models produced results very similar to those from the overall confirmatory factor analysis seen in Figure 2 suggesting that the presented results are fairly robust.

Model Using Maximum Likelihood Estimation With Robust Standard Errors (MLR)

Given potential issues with non-normal distributions with some of the measures we used maximum likelihood estimation with robust standard errors (MLR). This is based on the Satorra–Bentler scaled chi-squared test which is robust to non-normality (Curran et al., 1996; Satorra & Bentler, 1994). Adjustments with the Satorra-Bentler test also leads to robust standard errors, p-values, and confidence intervals. Therefore, we tested a version of the main confirmatory factor analysis using the Satorra-Bentler scaled chi-squared test. As shown in Table A2, the fit of the model was acceptable with factor loadings of the measures and the latent correlation between WMC and AC being very similar to the main confirmatory factor analysis.

Model Using Transformed Measures

Another way of dealing with non-normal data is to transform the skewed measures. Therefore, we transformed the accuracy variables (with an acrsine transformation) and the psychomotor vigilance task (with a log transformation). This resulted in overall more normal distributions for the measures. As shown in Table A2, the fit of the model was acceptable with factor loadings of the measures and the latent correlation between WMC and AC being very similar to the main confirmatory factor analysis.

Table A3

Estimated Power for Factor Loadings and Factor Correlations for the Simulated Models in Table 6

Construct/measure	180 Sim	120 Sim	360 Sim	Acc Sim	Comp Sim
WMC					
Ospan	.98	.94	1.00	1.00	.99
Symspan	.98	.94	1.00	1.00	.99
Rspan	.98	.94	1.00	1.00	.98
AC					
Anti	.88	.73	.99	.99	.96
Stroop	.43	.31	.71	.68	.50
Flanker	.81	.64	.98	.98	.93
WMC-AC *r*	.88	.71	.99	.99	.95

Note. Ospan = operation span; Symspan = symmetry span; Rspan = reading span; Anti = antisaccade.

Model Excluding Potential Multivariate Outliers

Given potential outliers could influence the results, we also checked for possible multivariate outliers and excluded participants with significant Mahalanobis's d2 values. This resulted in the removal of data for 34 participants. We then reran the model with these participants excluded. As shown in Table A2, the fit of the model was acceptable with factor loadings of the measures and the latent correlation between WMC and AC being very similar to the main confirmatory factor analysis. Thus, excluding potential outliers resulted in very similar overall results.

Estimates of Power for the Simulated Models in Table 6

We also examined power for the different simulated models seen in Table 6 with Wang and Rhemtulla's (in press) pwrSEM app. We specified the factor loadings and factor correlation based on the loadings and correlations in the simulated models with 1000 samples per model. These results suggested that similar to the overall simulation results, that with smaller sample sizes there was generally insufficient power to reliably detect loadings of the AC measures onto the AC construct. As sample size increased (or accuracy or the composite variables were used), power tended to increase (although power to detect the Stroop loadings were still insufficient) to more sufficient levels. Increasing the *N* up to 475 resulted in sufficient power (.81) for the Stroop loadings.

Received December 6, 2019
Revision received September 17, 2020
Accepted September 21, 2020 ■