

Note: This manuscript has been submitted for publication and is currently under review.

Nature and Measurement of Attention Control

Alexander P. Burgoyne¹, Jason S. Tsukahara¹, Cody A. Mashburn¹,
Richard Pak², & Randall W. Engle¹

¹Georgia Institute of Technology

²Clemson University

Funding. This work was supported by Office of Naval Research Grants N00173-20-2-C003 and N00173-20-P-0135 to Randall W. Engle.

Word count: 21,418

Abstract

Individual differences in the ability to control attention are correlated with a wide range of important outcomes, from academic achievement and job performance to health behaviors and emotion regulation. Nevertheless, the theoretical nature of attention control as a cognitive construct has been the subject of heated debate, spurred on by psychometric issues that have stymied efforts to reliably measure differences in the ability to control attention. For theory to advance, our measures must improve. We introduce three efficient, reliable, and valid tests of attention control that each take less than three minutes to administer: Stroop Squared, Flanker Squared, and Simon Squared. Two studies (online and in-lab) comprising more than 600 participants demonstrate that the three “Squared” tasks have great internal consistency (avg. = .95) and test-retest reliability across sessions (avg. $r = .67$). Latent variable analyses revealed that the Squared tasks loaded highly on a common factor, which was strongly correlated with an attention control factor based on established measures (avg. $r = .81$). Moreover, attention control correlated strongly with fluid intelligence, working memory capacity, and processing speed, and helped explain their covariation. We found that the Squared attention control tasks accounted for 75% of the variance in multitasking ability at the latent level, and that fluid intelligence, attention control, and processing speed *fully* accounted for individual differences in multitasking ability. Our results suggest that Stroop Squared, Flanker Squared, and Simon Squared are reliable and valid indicators of attention control. The tasks are freely available online.

Individual differences in the ability to control attention are correlated with a wide range of important outcomes, from cognitive task performance (Burgoyne et al., 2021; Conway et al., 2002; Draheim et al., 2021, 2022; Engle et al., 1999; Martin et al., 2020a; McVay & Kane, 2012) and academic achievement (Ahmed et al., 2019; Best et al., 2011) to health behaviors (Allan et al., 2016; Hall et al., 2008) and emotion regulation (Baumeister et al., 2007; Schmeichel & Demaree, 2010; Zelazo & Cunningham, 2007). As such, considerable time and effort has been invested in research on the nature of individual differences in attention control and their measurement (Lezak, 1982; McCabe et al., 2010; Willoughby et al., 2011; Zelazo et al., 2013).

Attention control refers to the domain-general ability to regulate information processing in service of goal-directed behavior (Burgoyne & Engle, 2020; Engle, 2002, 2018; Shipstead et al., 2016). More specifically, attention control allows us to maintain focus on task-relevant information while resisting distraction and interference by external events and internal thoughts. We have argued that the ability to control attention is important for a wide range of cognitive tasks, helping to explain why measures of cognitive abilities correlate positively with one another (Burgoyne et al., 2022; Kovacs & Conway, 2016). Attention control supports two distinct but complementary functions in our theoretical framework: maintenance and disengagement (Burgoyne & Engle, 2020; Shipstead et al., 2016). Whereas maintenance refers to keeping track of goal-relevant information, disengagement refers to removing irrelevant (or no-longer-relevant) information from active processing and tagging it for non-retrieval. Both functions require “attention control,” although they can also be modeled as separate but correlated latent factors (see Martin et al., 2020b).

Within the broader literature, attention control has also been referred to using terms such as “cognitive control” (Botvinick et al., 2001), “executive functions” (Diamond, 2013; Miyake et

al., 2000), “executive attention” (Engle, 2002), and the “central executive” (Baddeley, 1996).

Given its many names, it should come as no surprise that there are also many theoretical accounts of attention control. In addition to our “executive attention” view, the Friedman-Miyake model of executive functions has been particularly influential (Miyake & Friedman, 2012). Specifically, in their model, a higher-order inhibition factor is theorized to account for the covariation between lower-order updating and shifting factors (Miyake & Friedman, 2012). Our interpretation of this result is that it is largely consistent with our theoretical framework; what Miyake and Friedman (2012) refer to as “inhibition” is subsumed by what we refer to as “attention control.”

The Executive Attention View of Attention Control

Interest in attention control as a cognitive construct has been driven in part by the strong relationship between *working memory capacity*, reflecting the ability to maintain and manipulate information amidst interference, and *fluid intelligence*, reflecting novel problem solving and reasoning ability, including the ability to disengage from previous solution attempts (Shipstead et al., 2016). Early on, researchers observed a very strong correlation between these two constructs at the latent level, leading some to suggest that fluid intelligence may reflect little more than working memory capacity (Kyllonen & Christal, 1990). Today, we know that fluid intelligence and working memory capacity are distinct (Ackerman et al., 2005; Kane et al., 2005; Oberauer et al., 2005). Nevertheless, an explanation for their strong correlation has been the subject of heated debate (see, e.g., Burgoyne et al., 2019; Kane & Engle, 2002; Salthouse & Pink, 2008; Wiley & Jarosz, 2012). Our research has attempted to explain this relationship by identifying cognitive mechanisms that are shared across tests of fluid intelligence and working memory capacity.

Over twenty years ago, Engle et al. (1999) argued that if working memory capacity reflects the interplay between short-term memory and executive attention, then it is the executive attention component that largely explains working memory capacity's relationships with other cognitive constructs, including fluid intelligence. By measuring working memory capacity, short-term memory, and fluid intelligence at the latent level, Engle et al. (1999) showed that it was not short-term storage that drove working memory capacity's relationship to fluid intelligence, but rather, the additional attentional processes demanded by complex span tests of working memory capacity that are not demanded by short-term memory tests. That is, working memory capacity tests require both storage *and* concurrent processing of information, and this additional cognitive processing is what appeared to largely account for the relationship between working memory capacity and fluid intelligence. Specifically, Engle et al. (1999) found that after accounting for short-term memory, working memory capacity still predicted individual differences in fluid intelligence, whereas after accounting for working memory capacity, short-term memory did not account for significant variance in fluid intelligence.

Following Engle et al. (1999), Conway et al. (2002) conducted another latent variable analysis, this time to determine whether processing speed played a role in the relationship between working memory capacity and fluid intelligence. Their analyses showed that even after controlling for processing speed and short-term memory, working memory capacity still had a significant relationship with fluid intelligence, whereas processing speed and short-term memory were not significantly related to fluid intelligence after accounting for working memory capacity. This reinforced Engle et al.'s (1999) findings by showing that speed of information processing, like short-term memory, was not the primary driver of the working memory capacity-fluid intelligence relationship. Again, the evidence suggested that it was the executive attention

component of the working memory system that was the underlying factor driving the relationship between working memory tests and higher-level and real-world cognitive tasks.

More recently, Unsworth et al. (2014) extended this work by examining the relative contributions of attention control, short-term storage capacity, and retrieval from secondary memory to fluid intelligence. Their analyses added nuance to the conclusions of Engle et al. (1999) and Conway et al. (2002) by suggesting that retrieval from secondary memory might also help explain the relationship between working memory capacity and fluid intelligence. In their model, attention control, short-term storage capacity, and retrieval from secondary memory fully accounted for the relationship between working memory capacity and fluid intelligence. Taken together, latent variable analyses have repeatedly shown that attention control (i.e., the executive attention component of the working memory system) plays an important role in explaining a significant portion of the relationship between working memory capacity, fluid intelligence, and myriad other cognitive tasks such as general sensory discrimination (Tsukahara et al., 2021).

That said, most latent variable studies supporting the executive attention view have used working memory tasks as a proxy for the executive attention component of working memory. To advance our understanding of the nature of attention control we need to directly measure it and then model it at the latent level. In this regard, our conclusions about attention control have been limited by the quality of the measures available to researchers.

The Challenge of Measuring Individual Differences in Attention Control

Reliably measuring individual differences in attention control has posed a challenge to researchers and created a considerable barrier to theory development and real-world application. Simply put, most tasks used to measure individual differences in attention control suffer from poor reliability (Hedge et al., 2018; Rouder & Haaf, 2019), with only a few notable exceptions,

such as the Antisaccade task (Hallett, 1978). Because unreliability attenuates (i.e., reduces) the observed relationship between measures (Lord & Novick, 2008), most measures of attention control correlate weakly with each other or with other measures that are hypothesized to tap controlled attention (Draheim et al., 2019; Hedge et al., 2018; Paap & Saawi, 2016), which can result in a fractionated latent structure (Friedman & Miyake, 2004). Low reliability can also lead researchers to accept the null hypothesis about relationships with attention control at the individual task and construct level if the researchers are so inclined (Rey-Mermet et al., 2018). However, that does not mean no relationship exists, just that the measurement is inadequate to observe it.

As is now well-documented (Draheim et al., 2019; Hedge et al., 2018; Rouder & Haaf, 2019), part of the reliability problem can be attributed to psychometrically unsound tasks that use *response time difference scores* as the outcome measure, including “classic” experimental paradigms such as the Stroop (Stroop, 1935), Flanker (Eriksen, 1974), and Simon tasks (Simon & Rudell, 1967) (see Figure 1). Although the Stroop, Flanker, and Simon paradigms are great tools for *experimental* psychology, they suffer from severe limitations when used as-is for the study of individual differences (i.e., *differential* psychology). This phenomenon, which has been referred to as “the reliability paradox” (Hedge et al., 2018, p. 1166), is a product of the minimal between-subjects variance in the experimental effect of conflict tasks. From an experimental perspective, a manipulation is effective (and reliable) when it generates a similar effect for all participants, but from an individual-differences perspective, there must be systematic differences in the effect across individuals for the magnitude of the effect to correlate with other theoretically relevant measures. Thus, tasks that are well-suited for experimental research are

NATURE AND MEASUREMENT OF ATTENTION CONTROL

often poorly suited for individual differences research because they rely on an unreliable reaction-time difference score.

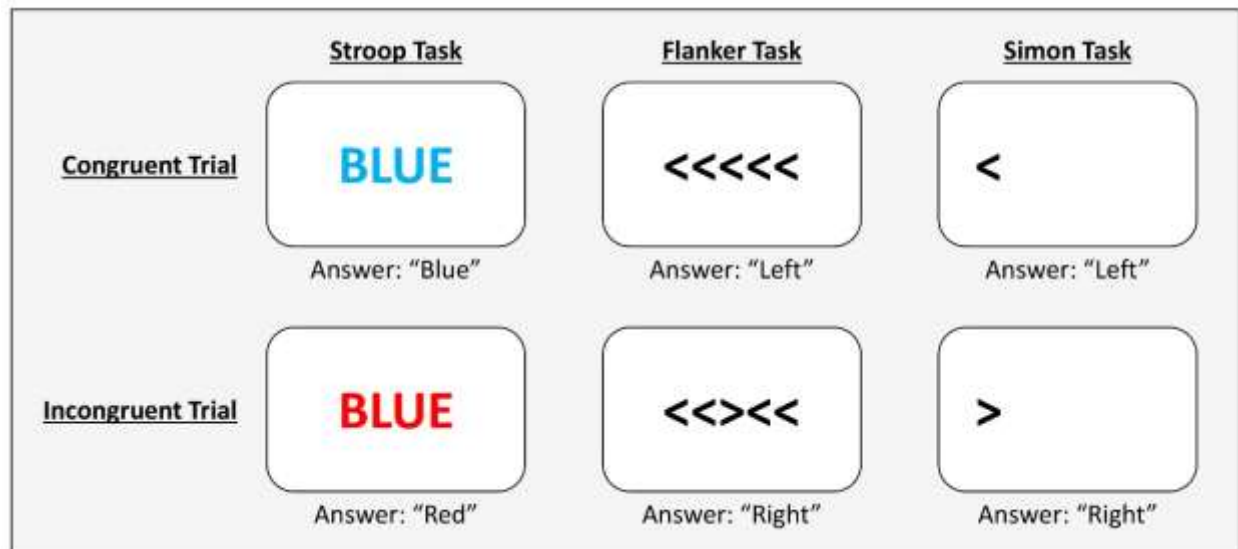


Figure 1. Examples of congruent and incongruent trials from the classic Stroop, Flanker, and Simon tasks. In the Stroop task, participants must indicate the color the word is printed in while disregarding the word's meaning. In the Flanker task, participants must indicate which direction the central arrow is pointing while disregarding the flanking arrows. In the Simon task, participants must indicate which direction the arrow is pointing while disregarding which side of the screen it appears on.

Consider the Stroop task. Participants must indicate the color a word is printed in, *not* the color the word refers to. Trials can be *congruent*, such as when the word "BLUE" is printed in blue ink, or *incongruent*, as when the word "BLUE" is printed in red ink. Incongruent trials demand the control of attention because participants must resolve conflict between the word's meaning and its color. By contrast, congruent trials require largely non-attentional processes because reading is a highly automated for most adults and there is no conflict between the stimulus's meaning and its color (MacLeod, 1991). The difference in response times on incongruent and congruent trials is thought to reflect attention control-related variance, and for this reason many tasks from the experimental psychology tradition such as the Stroop, Flanker,

and Simon paradigms use response time difference scores between congruent and incongruent trial conditions as the outcome measure.

The “subtraction method” (Donders, 1868) has been a valuable tool for experimental researchers (Chiou & Spreng, 1996). Studies consistently show that participants are slower to respond to incongruent trials than congruent trials, suggesting that incongruent trials are more cognitively demanding than congruent trials (MacLeod, 1991). Or, as Haaf and Rouder (2017) recently put it, “everybody Stroops” (p. 779). That said, psychometricians have cautioned against the use of difference scores in individual differences research for decades because of their unreliability at the level of the participant, and subsequently poor validity (Ackerman & Hambrick, 2020; Cronbach & Furby, 1970; Draheim et al., 2016; 2019; Friedman & Miyake, 2004; Hedge et al., 2018).

Difference scores are less reliable than their component scores (e.g., performance measures on congruent and incongruent trials) because subtraction removes the shared—and therefore reliable—variance of the component scores while preserving the error variance (i.e., noise). As the correlation between performance on congruent and incongruent trials increases, the reliability of the resulting difference score decreases, and is exacerbated by the unreliability of the component scores (see Figure 2).

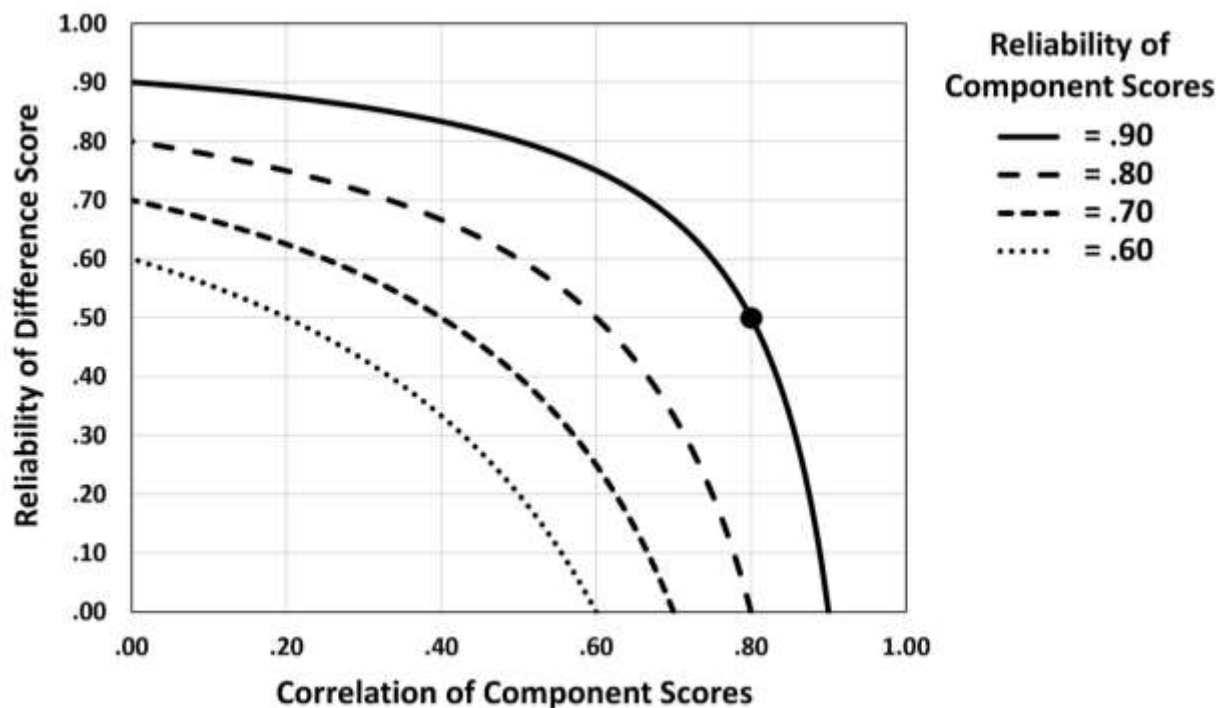


Figure 2. The reliability of a difference score (Y-axis) decreases as the correlation between the component scores (i.e., performance on congruent trials and incongruent trials) increases (X-axis). Each line represents the reliability of the difference score when the reliability of the component scores is set to .60, .70, .80, or .90. For a typical attention control task, one might find correlations between component scores to be around .80, and the reliability of each component score to be around .90, leading to a difference score reliability of .50, depicted by a black circle in the figure. Note that if the reliability of each component score simply decreased from .90 to .80 while the correlation between them remained .80, the resulting difference score would have a reliability of zero. Figure adapted from Draheim et al. (2019).

Using results from our lab as an example (Draheim et al., 2021), measures of performance on congruent and incongruent Stroop trials are typically strongly correlated (around $r = .80$) and have good reliability (around $\alpha = .90$). Given these values, the reliability of the resulting difference score is only $\alpha = .50$ (see Figure 2, above), meaning that only 25% of its variance reflects the construct of interest! The upshot is that given two difference score measures with reliabilities of $\alpha = .50$ —for example, Stroop performance and Flanker performance—the

observed correlation between them will be *half the magnitude* of the true correlation (i.e., the correlation if the measures were perfectly reliable) (see Figure 3, below).

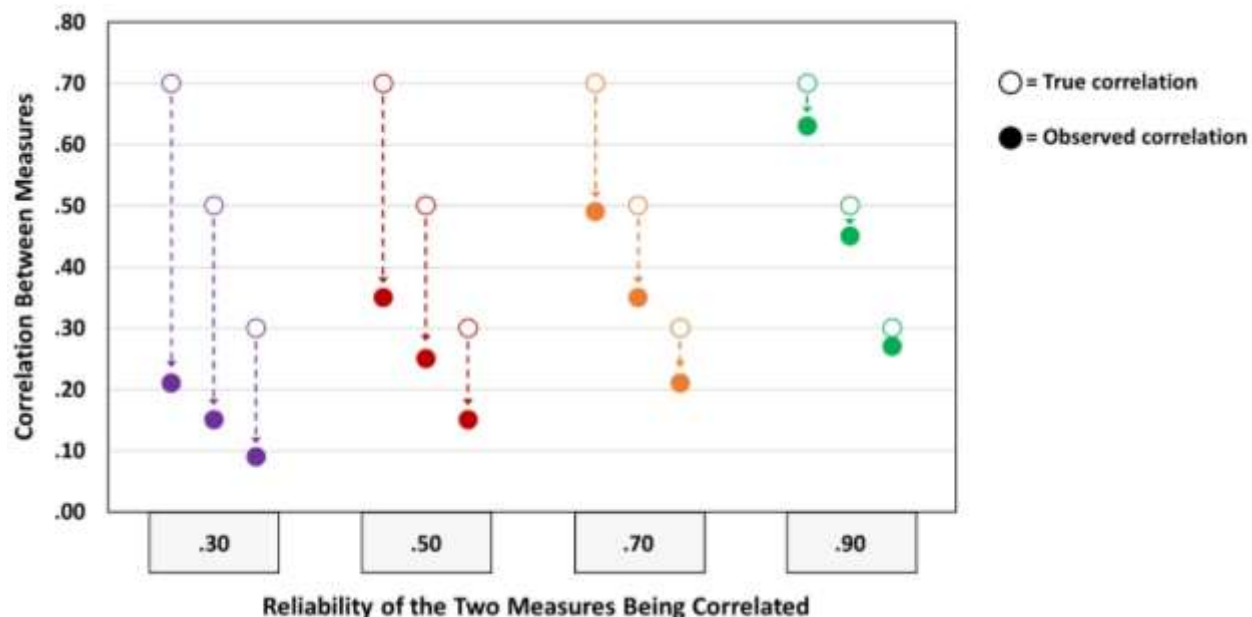


Figure 3. The attenuating effect of unreliability on the observed correlation between measures. Open circles depict the true correlation, reflecting the relationship between two measures given perfect reliability. Filled circles depict the observed correlation between two measures if both measures have reliabilities of .30, .50, .70 or .90.

Thus, three “classic” attention tasks used in the experimental psychology tradition, the Stroop, Flanker, and Simon paradigms, all suffer from an unreliability problem that has stymied the study of individual differences in attention control. We think the field would benefit from improved tests of attention control with better psychometric properties and that is the focus of this paper.

Previous Solution Attempts by Our Laboratory

To this end, our laboratory recently developed new attention control tests that avoided the use of response time difference scores (Draheim et al., 2021). For example, we modified the classic Stroop and Flanker tasks to use an adaptive response deadline. Participants were

challenged to respond to each item within a given time limit. If they responded accurately before the response deadline, the deadline for each trial became shorter, requiring quicker responses. If they could not respond accurately in time, the response deadline became longer, allowing slower responses. This thresholding approach converged on the rate at which participants could maintain a critical accuracy rate (for instance, .75), which was held constant across participants. The measure of performance was the duration of the response deadline at the conclusion of the task, with shorter deadlines indicating better performance and greater attention control.

The Stroop and Flanker tasks that used adaptive response deadlines had better test-retest reliability than the classic tasks they were modeled after, but still left room for improvement. For example, the test-retest reliability of the Flanker Adaptive Deadline task was $r = .54$ after removing outliers, far better than that of the classic Flanker task ($r = .23$). The Stroop Adaptive Deadline task had a test-retest reliability of $r = .67$, which was slightly better than that of the classic Stroop task ($r = .46$). These test-retest reliability estimates might have been higher if the time between testing sessions was reduced; on average, the time between testing sessions was six months. Draheim et al. (2021) did not compute an internal consistency reliability coefficient (e.g., Cronbach's alpha or split-half reliability) for the adaptive deadline tasks because task parameters (i.e., response deadlines) change over the administration of the test, rendering internal consistency estimates difficult to interpret in the usual manner.

One issue with the adaptive deadline tasks is that, although they were programmed to converge on the same critical accuracy rate for all participants (75%), in practice, accuracy rates varied widely. For example, for the Flanker Adaptive Deadline task, the average accuracy rate was 87.6% ($SD = 3.4\%$), and the range was quite large (69.4% - 95.1%). One potential explanation for this result is that the thresholding procedure assumes that the participant will

maintain the same ability level for the duration of the task, however, effort, motivation, attention, and fatigue can fluctuate over the testing session. Thus, if a person loses motivation midway through the task, their accuracy rate will drop, and the converged upon difficulty threshold will not reflect their “true” maximum ability level.

Another approach [Draheim et al. \(2021\)](#) used to develop new attention control tests was to create new tasks that demanded controlled attention but relied on accuracy and made response times irrelevant to performance. For example, the novel Sustained Attention to Cue task challenged participants to fixate on a circle that remained at a particular spatial location on the computer monitor. After a variable delay of two to 12 seconds, a distractor asterisk would flicker somewhere else on the screen, and then a letter would briefly appear at the spatial location cued by the circle, followed by a visual mask. Participants needed to sustain focus on the spatial location of the circle and inhibit an eye movement to the flickering asterisk in order to detect the briefly presented letter.

On its face, the Sustained Attention to Cue task shares similarities with the Antisaccade task, a “gold-standard” measure of attention control; both require inhibiting an eye movement to a salient distractor stimulus (i.e., a flickering asterisk) to detect a briefly presented letter at a different location. They differ in that, in the Antisaccade task, the flickering asterisk serves as a spatial cue, a temporal cue, and a distractor, whereas in the Sustained Attention to Cue task, the flickering asterisk is only a temporal cue and a distractor (i.e., it is not a spatial cue). By way of example, in the Antisaccade task, participants do not know *when* or in which of two locations the target letter will appear until the onset of the flickering asterisk; they must register the location of the flickering asterisk and immediately look the opposite direction to detect the letter. In the Sustained Attention to Cue task, participants know where the target letter will appear before

seeing the asterisk, because it is cued by a circle. However, they do not know *when* the target letter will appear; this is cued by the flickering asterisk. One related issue with the Sustained Attention to Cue task is that because the circle cue remained on the screen for the duration of the wait interval, attention could potentially drift away from the cued spatial location and then return to the cued location without much loss in performance, because the circle would remind them where to fixate after suffering from an attentional lapse. (We note that this issue has been fixed in the revised version of the task we used here; see the Method).

The internal consistency reliability of the Sustained Attention to Cue task ($\alpha = .93$) rivaled that of the Antisaccade ($\alpha = .92$)—both values are considered excellent. By comparison, the test-retest reliability of the Sustained Attention to Cue task was $r = .63$, slightly lower than that of the Antisaccade ($r = .73$) but still good. Thus, from a psychometric perspective the Sustained Attention to Cue task performed well. Nevertheless, it could be argued that the original version of the Sustained Attention to Cue task was *too* similar to the Antisaccade task, because both tasks share method-specific variance. We addressed this limitation of the Sustained Attention to Cue task in the revised version, which we use in the present studies.

Another approach to improving the measurement of attention control was developed by Martin et al. (2021) and adopted by Draheim et al. (2021): using Selective Visual Arrays as a measure of attention control. In the original Visual Arrays task (i.e., change detection task; Luck & Vogel, 1993), participants are challenged to remember a briefly presented array of colored squares. After a short delay, a second array of colored squares appears and participants must indicate whether anything in the array (or a particular item in the array) changed. In the *Selective* version of the Visual Arrays task, participants are pre-cued to memorize only a subset of the stimuli in the first array, for instance, either the *red* or *blue* rectangles. They are then shown two

arrays, the first consisting of red and blue rectangles, and the second consisting of just the cued color rectangles, with a delay in between them. Participants are asked whether the orientation of one of the cued-color items changed.

As thoroughly detailed by Martin et al. (2021), the *Non-Selective* Visual Arrays task loads more highly on a latent factor representing working memory capacity than it does on an attention control factor. This accords with the traditional view of Visual Arrays as a measure of visual working memory capacity (Luck & Vogel, 1993). *Selective* Visual Arrays, however, appears to have split loading on working memory capacity and attention control, likely due to the attentional filtering demand posed by the pre-cue. Attentional filtering is crucial, because if a participant cannot selectively attend to the cued subset of items and block representation of the uncued items, then the memory demand of the array is doubled, because the participant must try to remember *all* the items instead (Fukuda et al., 2015). Although Martin et al. (2021) make a compelling case for *Selective* Visual Arrays as a measure of attention control, they note that this view has generated considerable pushback from reviewers who still view visual arrays (selective or otherwise) as a measure of visual working memory capacity.

Overall, the four best attention control tasks to emerge from Draheim et al. (2021) were the Antisaccade, Sustained Attention to Cue (i.e., SACT), Flanker Adaptive Deadline (i.e., FlankerDL), and Selective Visual Arrays. These tasks were more reliable than the classic Stroop and Flanker tasks, demonstrated larger average correlations with other attention control tasks, and loaded more highly on a common attention control factor.

As we stated, having a theory of attention control depends on understanding attention control at the construct level, and that, in turn, depends on having reliable and valid measures of the construct. There remains room for improvement in the measurement of attention control, as

we have detailed in the preceding paragraphs. Moreover, from a practical perspective, the four best tasks from Draheim et al. (2021) require approximately one hour of testing time, which significantly hampers researchers' ability to measure other psychological constructs in addition to attention control within a single session of data collection. Furthermore, lengthy testing time reduces the likelihood that a measure will be used in studies directed at transitioning from basic to applied research.

Goals of The Present Studies

In the present studies, we build on our laboratory's previous work by showcasing three efficient, reliable, and valid measures of attention control that each take less than three minutes to administer: Stroop Squared, Flanker Squared, and Simon Squared. All three tasks are "new takes" on their classic experimental paradigm counterparts that avoid the use of response time difference scores, adaptive thresholding, rapid visual presentation, and lengthy testing time. Furthermore, the tasks are gamified, featuring a self-scoring points system, timer, sound effects, and a 'point-and-click' interface. We tested these tasks alongside the best attention control tasks to emerge from Draheim et al. (2021) and measures of other cognitive constructs in an online study (Study 1) and an in-laboratory study (Study 2).

Our analyses examine the internal consistency reliability, test-retest reliability, convergent validity, discriminant validity, and predictive validity of the three "Squared" tests of attention control (i.e., Stroop Squared, Flanker Squared, and Simon Squared; see the next section for descriptions of each task). We estimate the tests' split-half internal consistency using Spearman-Brown's prophecy formula, and, for Study 2, we also estimate test-retest-retest reliability and practice effects over three testing sessions: two in the laboratory and one online. We estimate convergent validity by examining correlations at the observed and latent level

between the three Squared tests of attention control and the best attention control tests to emerge from Draheim et al. (2021). Finally, we examine predictive validity by estimating the relationship between performance on the three Squared tests of attention control and performance on a battery of fluid intelligence, working memory capacity, processing speed, and multitasking paradigms.

We were particularly interested in determining whether the three Squared tests of attention control could account for the positive correlations observed among cognitive ability measures (i.e., the *positive manifold*; Spearman, 1904) to a similar degree to the best attention control tasks to emerge from Draheim et al. (2021). Although studies have shown that attention control can partly explain the covariance between constructs such as working memory capacity, fluid intelligence, and sensory discrimination ability (Engle et al., 1999; Conway et al., 2002; Unsworth et al., 2014; Tsukahara et al., 2020; Draheim et al., 2021; Burgoyne et al., 2022), whether a similar pattern of results will be obtained using the new Squared tests remains an open question. Thus, throughout the Results sections we report sets of latent variable analyses in which the attention control factor is defined by either the new Squared tests of attention control or the best tests to emerge from Draheim et al. (2021).

We also used latent variable modeling to investigate whether attention control or processing speed plays a more fundamental role in explaining the relationships between cognitive abilities. The debate over the importance of processing speed arises from an increase in the use of drift diffusion modeling, which decomposes accuracy and reaction time data in two-alternative forced choice tasks to identify parameters presumed to reflect cognitive processes involved in decision making (Ratcliff & Rouder, 2000). Drift diffusion modeling assumes that evidence accumulates over time towards a response threshold (or boundary), and once this

boundary is reached, a response is initiated. Using drift diffusion modeling, some researchers have argued that drift rate, or speed of evidence accumulation, reflects processing speed (Lerche et al., 2020), and have shown that drift rate is correlated across classic conflict tasks used to measure attention control. Although this work would appear to suggest that what is reliably measured by conflict tasks is drift rate (among other things), we take issue with the interpretation of these results that equates drift rate to processing speed without considering where attention control fits into the model. This is because evidence indicates that drift rate is strongly influenced by the focus of attention. For example, Kofler et al. (2020) found that instructing participants to simultaneously complete a secondary task while making judgments in a two-alternative forced choice paradigm significantly lowered participants' drift rate, which indicates that what we pay attention to (and also, our ability to focus attention on task-relevant information) influences the rate of evidence accumulation in drift diffusion models. In other words, even in the absence of a secondary task, trial-to-trial lapses in attention will result in some people having a faster average drift rate than others simply because they are better able to maintain focus on the task at hand. Stated differently, we think that attention control influences drift rate, and therefore may be a more fundamental cognitive construct when it comes to explaining variance (and covariance) in complex task performance.

Finally, we wanted to test whether individual differences in attention control (and in particular, performance on the Squared tasks) could account for individual differences in multitasking ability. Multitasking refers to the process by which individuals juggle multiple subtasks or information processing demands concurrently (or in an interleaved fashion) in service of a goal. As such, multitasking is a complex cognitive activity, designed to be a proxy for real-world work situations. The sub-tasks draw on many executive functions, such as the

ability to maintain overarching goals, switch between subtasks, disengage from no-longer relevant information, avoid mind wandering, distractions and interference, and also strategically allocate resources (e.g., time, effort, attention) to maximize performance. However, multitasking also requires problem solving and rapidly responding to goal-relevant stimuli. It follows that multitasking likely requires the interplay between not only attention control but also other cognitive abilities such as fluid intelligence, processing speed, and working memory, necessitating work that sheds light on the amount of unique variance that each of these constructs captures in multitasking performance.

Indeed, the evidence suggests that individual differences in attention control and other cognitive abilities play a role in multitasking. For example, [Martin et al. \(2020a\)](#) examined the relative contributions of attention control, fluid intelligence, and performance on the Armed Services Vocational Aptitude Battery to multitasking ability at the latent level. On its own, the Armed Services Vocational Aptitude Battery—a standardized test used by the U.S. military for personnel selection—accounted for a majority of the variance in multitasking performance. When adding attention control and fluid intelligence to this model and allowing the predictors to correlate, however, attention control and fluid intelligence fully accounted for the predictive validity of the Armed Services Vocational Aptitude Battery. That is, the Armed Services Vocational Aptitude Battery was no longer a significant predictor, whereas attention control and fluid intelligence had substantial and similar-in-magnitude predictive paths to multitasking ability. Thus, attention control and fluid intelligence appear to capture significant unique variance in multitasking ability at the latent level, above and beyond one another. [Martin et al. \(2020a\)](#) also explored whether processing speed would account for variance in multitasking ability that was previously attributed to attention control. Instead, they found the opposite:

including processing speed did not add significant predictive value to the model, whereas the path from attention control to multitasking ability remained statistically significant and similar in magnitude. Whether the inclusion of working memory capacity would alter this pattern of results is a question we explore in the present work in Study 2. Given evidence suggesting that attention control is the primary “active ingredient” in measures of working memory capacity (e.g., Engle et al., 1999), we predicted that working memory capacity would not contribute significantly to the model once attention control was accounted for.

Introducing the Three Squared Tests of Attention Control

In this section, we introduce the Stroop Squared, Flanker Squared, and Simon Squared tasks. These tasks were designed to add an additional level of conflict to each of the traditional conflict paradigms—hence the use of “Squared.” They were also designed to have a short administration time: participants are given 90 seconds to earn as many points as possible. They earn one point for each correct response and lose one point for each incorrect response. The design of the tasks was inspired by the Double Trouble task from the Cambridge Brain Sciences Neurocognitive Battery (<https://www.cambridgebrainsciences.com/science/tasks/double-trouble>).

Stroop Squared

In Stroop Squared (Figure 4), participants are shown a target stimulus in the center of the screen with two response options below it. The target stimulus (“RED” or “BLUE” displayed in red or blue colors) follows the typical Stroop paradigm (see Figure 1) where a response must be made to the display color and not the semantic meaning of the word. However, what must be attended to in the response options is the meaning of the word—not the display color. The participant’s task is to select the response option with the word meaning that matches the display

color of the target stimulus. For example, if the target stimulus is the word “RED” appearing with a blue display color, the participant must select the response option that says the word “BLUE,” regardless of the response option’s display color. Thus, the challenge is for participants to pay attention to the display color of the target stimulus and the semantic meaning of the response options. Conversely, they must try to ignore the semantic meaning of the target stimulus and ignore the display color of the response options.



Figure 4. Stroop Squared. The participant’s task is to select the response option with the word meaning that matches the display color of the target stimulus. In the above example, the target stimulus is the word “RED” appearing with a blue display color, so the participant must select the response option that says the word “BLUE” (i.e., the one on the right).

Flanker Squared

In Flanker Squared (Figure 5), participants are shown a target stimulus and two response options. The target stimulus and response options are flanker items consisting of 5 arrows (e.g., > > < > >). The participant’s task is to select the response option with a central arrow that points in the same direction as the flanking arrows in the target stimulus. For example, given the following target stimulus (e.g., < < > < <), the participant must select the response option with a central arrow pointing to the left (e.g., > > < > >). Thus, the challenge is for participants to pay attention to the flanking arrows of the target stimulus and the central arrow of the response options.

Conversely, they must try to ignore the center arrow of the target stimulus and also ignore the flanking arrows of the response options.

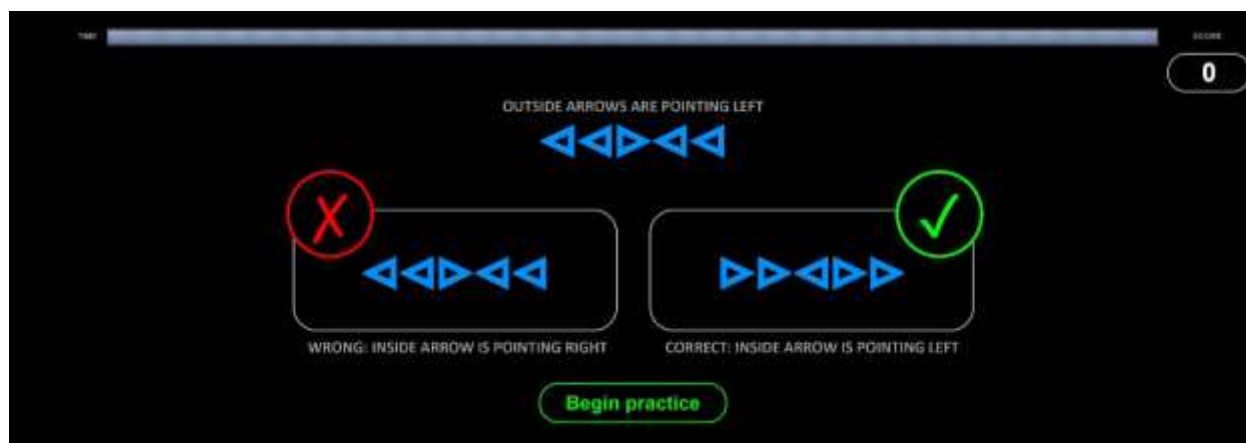


Figure 5. Flanker Squared. The participant’s task is to select the response option with a central arrow that points in the same direction as the flanking arrows in the target stimulus. In the above example, the target stimulus has flanking arrows pointing left, so the participant must select the response option which has a central arrow pointing left.

Simon Squared

In Simon Squared (Figure 6), participants are shown a target stimulus and two response options. The target stimulus is an arrow and the response options are the words “RIGHT” and “LEFT.” The participant’s task is to select the response option that states the direction that the arrow is pointing. For example, if the target stimulus is an arrow pointing left, the participant must select the response option that says the word “LEFT.” Complicating matters, the target stimulus arrow and response options can appear on either side of the computer screen with equal probability. Thus, the challenge is for participants to pay attention to the direction that the target stimulus arrow is pointing and the meaning of the response options. Conversely, they must try to ignore the side of the screen that the target stimulus arrow and response options appear on.

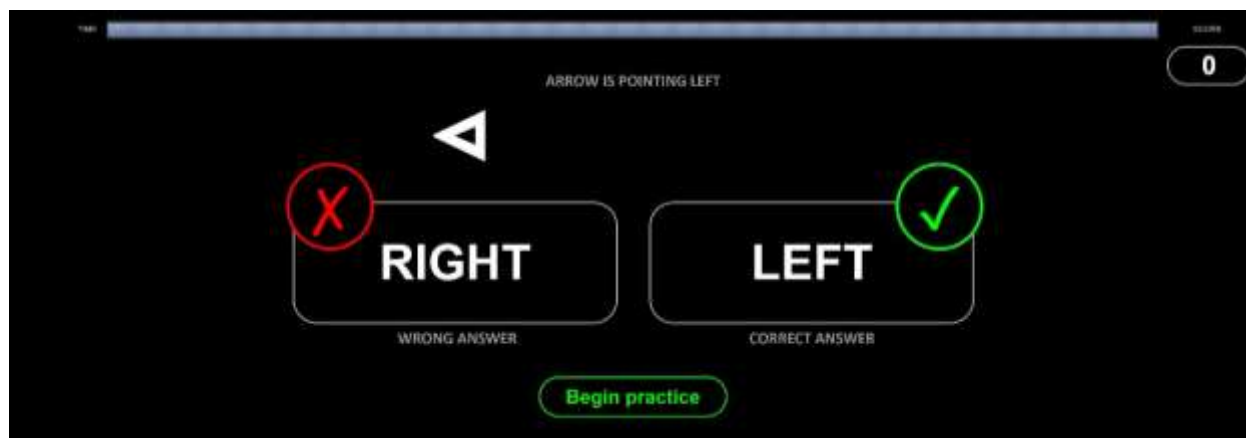


Figure 6. Simon Squared: The participant’s task is to select the response option that states the direction that the arrow is pointing. In the above example, the arrow is pointing left, so the participant must select the response option that says “LEFT”.

Trial Types in the Squared Tasks

In each of the Squared tasks, there are four trial types that are sampled with equal probability (see Figure 7). Trial types are defined by whether the target stimulus and response options are “congruent,” meaning the word’s semantic meaning and display color match (e.g., “RED” in red color in Stroop Squared), or “incongruent,” meaning the word’s semantic meaning and display color do not match (e.g., “RED” in blue color in Stroop Squared). The four trial types are *fully congruent*: the target stimulus and response options are all congruent; *fully incongruent*: the target stimulus and response options are all incongruent; *stimulus congruent, response options incongruent*: the target stimulus is congruent while the response options are incongruent; and *stimulus incongruent, response options congruent*: the target stimulus is incongruent while the response options are congruent.

In Studies 1 and 2, we explored whether there were any theoretically important differences across trial types in terms of performance or correlations with other cognitive constructs. One prediction was that fully congruent trials would be the easiest for participants because they require the least amount of conflict resolution and goal maintenance: participants

can match any stimulus attribute to a response option attribute to obtain the correct answer. Conversely, we expected that fully incongruent trials would be the most difficult, because they require the most amount of conflict resolution and goal maintenance. We did not have specific predictions regarding differences between the two types of partially incongruent trials, but anticipated that these trials would be moderately difficult for participants and demand conflict resolution and goal maintenance more than fully congruent trials but less than fully incongruent trials.

It seemed plausible that performance on fully incongruent trials might correlate more strongly with other measures of attention control than performance on congruent trials, given the difference in the amount of conflict resolution that must occur to successfully respond to each trial type. However, because all trial types were intermixed, with a superordinate goal carrying through the entirety of the task and being constantly reinforced (three-quarters of the trials involved navigating some amount of incongruency), it is possible that no differences will emerge in the correlations between performance on each trial type and attention control. Research on traditional conflict tasks has shown that the ratio of congruent to incongruent trial types affects correlations between performance and other cognitive abilities, such as working memory capacity (Hutchison, 2011; Kane & Engle, 2003); when incongruent trials are less frequent, performance is more strongly correlated with cognitive ability. Thus, we conducted analyses by trial type on a purely exploratory basis, as it would require a separate experiment to manipulate the ratio of different trial types and examine the consequences of doing so on correlations with cognitive ability.




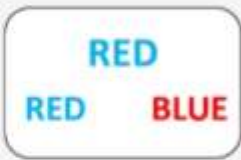
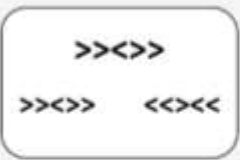







	<u>Stroop Squared</u>	<u>Flanker Squared</u>	<u>Simon Squared</u>
<u>Fully Congruent</u>			
<u>Fully Incongruent</u>			
<u>Stimulus Congruent, Response Options Incongruent</u>			
<u>Stimulus Incongruent, Response Options Congruent</u>			

Figure 7. Examples of the four trial types in the three Squared tests of attention control. The correct answer is the response option on the right for all example trials shown above. In Stroop Squared, the participant must select the response option with the word meaning that matches the display color of the target stimulus. In Flanker Squared, they must select the response option with a central arrow that points in the same direction as the flanking arrows in the target stimulus. In Simon Squared, they must select the response option that states the direction that the arrow is pointing.

Study 1: Mechanical Turk and Prolific

We first investigated the reliability and validity of the three Squared tests of attention control using an online sample of participants recruited through Amazon’s Mechanical Turk (MTurk) platform and Prolific.

Method

Participants

Our initial sample consisted of 375 participants recruited through MTurk and Prolific. Monte-Carlo simulations suggest that for stable estimates of correlations, sample sizes should approach 250 (Schönbrodt & Perugini, 2013). Our recruitment filters required participants to be ages 18-35, based in the United States, and, for MTurk, to have a work (i.e., HIT) approval rating greater than 92%. Additionally, our inclusion criteria stipulated that participants must be native English speakers with normal or corrected-to-normal vision, must not have had a seizure, and must have a Windows personal computer with internet access. All participants provided informed consent.

Procedure

This was an online study in which participants completed computerized tests of cognitive ability on their personal computers at their own pace. Almost all participants completed the study the same day that they began it. The tasks took around 2 to 2.5 hours to complete. The tasks were programmed using E-Prime Go (Psychology Software Tools, Pittsburgh, PA) and distributed to participants using a Qualtrics survey. Participants entered their worker ID into the survey, completed a PC check to ensure their computer was compatible with E-Prime Go, and then were given a link to download the tasks. Participants completed the tasks locally on their computer and the data files were automatically uploaded to our E-Prime Go dashboard as they completed each task. Participants entered a code into MTurk or Prolific to signify that they had completed the study, which was verified by the first author. Participants were paid \$30 for completing the study or a majority of the study's tasks. The task order was as follows: Demographics, Stroop Squared, Flanker Squared, Simon Squared, Antisaccade, Raven's Advanced Progressive Matrices, Advanced Symmetry Span, FlankerDL, Letter Sets, Advanced Rotation Span, SACT, Number Series, Selective Visual Arrays, Mental Counters.

Demographics

Participants were asked to report their age, gender, and ethnicity. They were asked whether English was the first language they learned and the age at which they learned it, and whether they were fluent in other languages. Participants were asked to report the highest level of education they had achieved as well as their annual household income. Participants were asked whether they had corrected vision, and also whether they had any conditions (e.g., illness, disability, medication use) that might affect their performance on cognitive tasks.

Attention Control

Stroop Squared. In Stroop Squared, participants must match the display color of the target stimulus with the semantic meaning of one of two response options . See Figure 4 above and the description of the task that accompanies it. Participants completed a 30-second practice phase followed by a 90-second test phase. Item-level feedback was provided on all trials. The measure of performance was the number of correct responses minus the number of incorrect responses.

For all of the Squared tasks, on the first screen of the task, participants were shown an example item (a “fully incongruent” trial) and were given instructions on how to complete the task. The correct response to the example item was indicated by a green checkmark and a description of why each response option was correct or incorrect. After reading the instructions, participants began a 30-second practice phase with feedback on every trial in the form of display text and a short auditory chime or buzzer. The participant’s current score was displayed in the top-right corner of the screen and the amount of time remaining was presented at the top of the screen. After 30 seconds of practice, participants were shown their score on the practice phase and taken back to the instructions screen for further review.

After reviewing the instructions again, the participant proceeded to the test phase by clicking the “start” button. A 3-second timer counted down, and then the test phase began. Participants were given 90 seconds to earn as many points as possible. Feedback was given on every trial in the same manner as during the practice phase. Participants could view their current score in the top corner of the screen and the amount of time remaining at the top center of the screen. After the 90-second test phase was completed, participants were told their final score and thanked for their participation.

Flanker Squared. In Flanker Squared, participants must match the direction of the flanking arrows of the target stimulus with the direction of the central arrow of one of two response options. See Figure 5 above and the description of the task that accompanies it. Participants completed a 30 s practice phase followed by a 90 s test phase. Item-level feedback was provided on all trials. The measure of performance was the number of correct responses minus the number of incorrect responses.

Simon Squared. In Simon Squared, participants must match the direction that a target stimulus arrow is pointing with the semantic meaning of one of two response options. See Figure 6 above and the complete description of the task that accompanies it. Participants completed a 30-second practice phase followed by a 90-second test phase. Item-level feedback was provided on all trials. The measure of performance was the number of correct responses minus the number of incorrect responses.

Antisaccade (Hallett, 1978; Hutchison, 2007). Participants identified a “Q” or “O” that appeared briefly on the opposite side of the screen as a distractor stimulus. After a central fixation cross appeared for 1000ms or 2000ms, an asterisk (*) flashed at 12.3° visual angle to the left or right of the central fixation for 100ms. Afterward, the letter “Q” or “O” was presented on

the opposite side at 12.3° visual angle of the central fixation for 100ms, immediately followed by a visual mask (##). Participants indicated whether the letter was a “Q” or an “O”. They completed 16 slow practice trials during which letter duration was set to 750ms, followed by 72 test trials. The task was scored based on accuracy as the proportion of correct responses.

Flanker Adaptive Deadline (FlankerDL; adapted from Draheim et al., 2021). The task was an arrow flanker task in which there was a target arrow in the center of the screen pointing either left or right along with two flanking arrows on both sides. The flanking arrows were either all pointing in the same direction as the central target (congruent trial) or all in the opposite direction (incongruent trial). There was a 2:1 ratio of congruent to incongruent trials with 96 incongruent trials and a total of 288 trials overall. The task was administered over 4 blocks of 72 trials each with an optional rest break between blocks. Practice trials were administered in different blocks, 18 standard flanker no deadline practice trials, and 18 non-adaptive response deadline practice trials.

An adaptive staircase procedure was used to estimate the subject’s response deadline threshold that converges around 60% accuracy. The adaptive procedure was based only on the incongruent trials. On each incongruent trial, if an incorrect response was made *or* the reaction time was longer than the response deadline then the response deadline increased (more time to respond) on the next trial. If a correct response was made *and* the reaction time was shorter than the response deadline, then the response deadline decreased (less time to respond) on the next trial. The response deadline started at a relatively easy level of 1.5 seconds. A 3:1 up-to-down ratio was used for the step sizes such that the step size (change in response deadline) for incorrect/too slow of trials was three times larger than the step size for correct/deadline met trials. The step size started at 240:80ms, decreased to 120:40ms after 17 incongruent trials,

decreased to 60:20ms after 33 incongruent trials, decreased to 30:10ms after 49 incongruent trials, decreased to 15:5ms after 65 incongruent trials, and finally settled at 9:3ms after 81 incongruent trials. Feedback was given in the form of an audio tone and the words “TOO SLOW! GO FASTER!” presented in red font when the response deadline was not met.

Importantly, this version of FlankerDL was adapted from Draheim et al. (2021) and differed in one significant way: In the previous version of the task, participants’ accuracy rate on *each block of 18 trials* determined whether the response deadline would increase or decrease. In this version of the task, participants’ accuracy rate on *each incongruent trial* determined whether the response deadline would increase or decrease. We made this change to the program to be more consistent with Kaernbach’s (1991) adaptive testing approach, which stipulates the use of trial-level information instead of block-level information when staircasing a task’s difficulty based on performance. Kaernbach’s (1991) guide was used when originally developing these tasks in our lab, however, this detail was overlooked in the previous version of the task and corrected in the version used here.

Sustained Attention to Cue (SACT; adapted from Draheim et al., 2021). The critical element in this task is the wait time interval in which attention must be sustained at a spatially cued location for a variable amount of time. After the variable wait time, a target letter is briefly presented and must be identified amidst a mix of other non-target letters. Each trial started with a central black fixation for 1 second followed by a 750ms interval in which the words “Get Ready!” were displayed at the to-be cued location along with an auditory beep. A circle cue was then displayed for approximately 500ms, and then was removed from the display during the wait time interval. The wait time lasted either 0 seconds or 2 – 12 seconds in 500ms intervals (e.g., 2, 2.5, 3, 3.5... seconds). After the variable wait time, a cloud array of letters was displayed at the

cued location for 250ms. The target letter was identifiable as the central letter in slightly darker font color. The target and non-target stimuli were B, P, or R's. The task had 3 blocks of 22 trials for a total of 66 trials without feedback. The task was scored as the proportion of correct responses.

The SACT task was also adapted from Draheim et al. (2021) and featured one major modification: In the previous version of the task, the fixation circle remained on the screen during the wait interval, so even if participants looked away from the target area they could re-attend to it again using the circle on the screen as a cue. In this version of the task, the fixation circle shrank in size to converge on the target area, and then disappeared for the duration of the wait interval. Thus, this version of the task challenged participants to remember the spatial location of the target area, because there was no circle on the screen to cue it during the wait interval.

Selective Visual Arrays (adapted from Luck & Vogel, 1997). After a central fixation of 1000ms, a cue word ("RED" or "BLUE") appeared instructing the participant to attend to either red or blue rectangles. Next, a target array of red and blue rectangles of different orientations (horizontal, left diagonal, right diagonal, and vertical) was presented for 250ms, followed by a blank screen for 900ms. The target array contained a set size of either 3 or 5 rectangles of each color. Next, a probe array with only the cued-color rectangles was presented, with one rectangle highlighted by a white dot. The orientation of the highlighted rectangle was either the same as it was in the target array, or different, with equal likelihood. The participant indicated with the keyboard whether the orientation of the highlighted rectangle had changed or stayed the same. There were 48 trials per array set size. Capacity scores (k) for each set size were calculated using

the single-probe correction (Cowan et al., 2005): $\text{set size} * (\text{hit rate} + \text{correction rejection rate} - 1)$. The task was scored as the mean k estimate across the two set sizes.

Fluid Intelligence

Raven's Advanced Progressive Matrices (Raven & Court, 1998). Participants were presented with 3×3 arrays of geometric patterns. Each array contained a missing item, and participants were to select the pattern that best completed the array. Participants were given 10 minutes to complete the 18 odd-numbered items from Raven's Advanced Progressive Matrices. The measure was the number correct.

Letter Sets (Ekstrom et al., 1976). Participants were presented with five sets of four letters arranged in a row, and were to select the set that did not follow the same pattern as the other four. For example, for the sets NLIK, PLIK, QLIK, THIK, and VLIK, the correct response is THIK because the other sets all contain L. Participants were given 10 minutes to complete 30 items. The measure was the number correct.

Number Series (Thurstone, 1938). Participants were presented with a series of numbers, and were to select which of four alternatives logically completed the series. Participants were given 5 minutes to complete 15 items. The measure was the number correct.

Working Memory Capacity

Advanced Symmetry Span (Unsworth et al., 2005). This task required subjects to remember a series of spatial locations in a 4×4 matrix presented in alternation with a pattern of squares which they had to decide whether the pattern was symmetrical on the vertical midline. On each trial, subjects were first presented with a 16×16 matrix of black and white squares and were required to decide whether the pattern was symmetric on the vertical midline. Followed by the symmetry judgment, a 4×4 matrix of squares with one square highlighted in red were

displayed. The location of the red-square was the to-be-remembered spatial location. This alternation continued until a variable set-size of spatial locations had been presented. Then, on the recall screen the same 4×4 matrix of squares was presented but with no squares highlighted in red. Subjects had to recall the spatial locations in the correct order by clicking the mouse on the appropriate squares in the matrix. There was a total of 12 trials (2 blocks of 6 trials), set-sizes ranged from 2-7, and each set-size occurred twice (once in each block). Scores on the advanced symmetry span task were calculated using the partial scoring method.

Advanced Rotation Span (Kane et al., 2004). This task required subjects to remember a series of directional arrows of varying size in alternation with a mental rotation task in which they had to mentally rotate and decide if a letter was mirror reversed or not. On each trial, subjects first solved a mental rotation problem followed by the presentation of a single arrow with a specific direction (8 possible directions; the four cardinal and four ordinal directions) and specific size (small or large). Both the direction and size of the arrow were the to-be-remembered features. This alternation continued until a variable set-size of arrows had been presented. Then, on the recall screen all possible arrow directions and sizes were presented. Subjects had to recall the direction and size of the arrows in the correct order by clicking the mouse on the appropriate arrow. There was a total of 12 trials (2 blocks of 6 trials), set-sizes ranged from 2-7, and each set-size occurred twice (once in each block). Scores on the advanced rotation span task were calculated using the partial scoring method.

Mental Counters (adapted from Alderton et al., 1997). This test challenged participants to keep track of three different values as they changed. Participants were presented with three lines in the center of the screen. On each trial, each line would begin with a value of five. Boxes would appear one at a time above or below the lines for 500 to 830 ms and then disappear, and

the participant's task was to add '1' to that line's value if a box appeared above the line and subtract '1' from that line's value if a box appeared below the line. After a series of boxes, the participant was asked to report the value for each of the three lines. There were five trials at set size five (e.g., five boxes appeared during the trial), 14 trials at set size six, and 13 trials at set size seven, for a total of 32 trials. The measure of performance was the partial score, reflecting the number of correctly reported values.

Data Preparation

We removed participants' scores on a task if they showed severely poor performance indicating they did not understand the instructions or were not performing the task as intended. Specifically, we computed chance-level performance on each task; any scores that were at or below chance-level performance were identified as problematic data points and set to missing. This procedure was applied to the three Squared tests of attention control, antisaccade, selective visual arrays, SACT, and FlankerDL. We did not remove datapoints representing sub-chance performance on the three fluid intelligence tests. For FlankerDL, we set trial-level performance to missing if the response time on that trial was less than 200ms, on the basis that these responses were too fast and likely represented mis-clicks. For the Advanced Span tasks, problematic data points were defined by chance-level performance or worse on the processing subtask. After removing 197 problematic data points (approximately 4% of the data), we performed a two-pass outlier exclusion procedure on all tasks. We removed data points that were more than 3.5 standard deviations worse than the sample mean two times, recomputing the sample mean and standard deviation each time. The outlier exclusion process removed 13 data points on the first pass and 13 data points on the second pass (< 1% of the data).

Modeling Approach and Fit Statistics

We used maximum likelihood estimation for all confirmatory factor analyses and structural equation models. We report multiple fit statistics: The χ^2 is an absolute fit index comparing the fit of the specified model to that of the observed covariance matrix. A significant χ^2 can indicate lack of fit, but is heavily influenced by sample size. In large samples, such as the one used in the present studies, even a slight deviation between the data and the model can lead to a significant χ^2 statistic. Therefore, we also report the comparative fit index (CFI) and Tucker-Lewis index (TLI), which compare the fit of the model to a null model in which the covariation between measures is set to zero, while adding penalties for additional parameters. For CFI and TLI, large values indicate better fit (i.e., $> .90$ or ideally, $> .95$). For the root mean square error of approximation (RMSEA) fit statistic, values less than $.05$ are considered great, while values less than $.10$ are considered only adequate. For the standardized root mean square residual (SRMR), which computes the standardized difference between the observed and predicted correlations, a value of less than $.08$ indicates good fit (Hu & Bentler, 1999).

Results

Demographic information is summarized in Table 1. The participants' average age was 27 ($SD = 5$) years old and a majority were female (53.7%). In terms of race/ethnicity, 62.6% of the sample identified as White, 10.3% identified as Black or African American, 7.5% identified as Asian or Pacific Islander, and the remainder selected "Other" or declined to respond. The majority of participants (86.2%) had attended at least some college.

Table 1
Demographic Information for Study 1

Demographic	Statistic
Age (years)	Mean = 27.28 $SD = 5.06$ Range = 18 – 35

Gender	Male = 43.7% Female = 53.7% Self-Identify/Other = 2.0% Transgender Male = 0.6%
At least some college?	Yes: 86.2% No: 13.8%
Ethnicity	White: 62.6% Black or African American: 10.3% Asian or Pacific Islander: 7.5% Other*: 19.2%

Note. Demographic information was unavailable for some participants, lowering the effective *N* to 348. *Other includes, Hispanic or Latino, Native American, and “Other”.

Descriptive statistics are presented in Table 2. Of the three Squared tests of attention control, participants earned the most points on Simon Squared ($M = 57.38$), followed by Stroop Squared ($M = 31.29$) and Flanker Squared ($M = 27.38$). Paired samples *t*-tests revealed that participants scored significantly higher on Simon Squared than on Stroop Squared ($t(297) = 33.14, p < .001$) and Flanker Squared ($t(290) = 39.80, p < .001$). This suggests that of the three Squared tests, Simon Squared may be the easiest for participants, whereas Stroop Squared and Flanker Squared may be more difficult.

The three Squared tests of attention control demonstrated excellent internal consistency reliability: Stroop Squared (.93; average number of trials = 42), Flanker Squared (.94; average number of trials = 37), Simon Squared (.97; average number of trials = 61). These split-half internal consistency estimates were computed by correlating performance on odd-numbered and even-numbered trials (because the total number of trials varied across participants) and using the Spearman-Brown correction. The reliability of the three Squared tests of attention control was as good or better than the reliability of the other attention control tests: Antisaccade (.87), FlankerDL (.89), SACT (.95), and Selective Visual Arrays (.58).

Table 2. *Descriptive statistics for Study 1.*

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Skew</i>	<i>Kurtosis</i>	<i>Reliability</i>	<i>Time</i>
Stroop Squared	311	31.29	14.64	-0.45	-0.44	.93 ^b	2 min
Flanker Squared	297	27.38	13.96	-0.14	-0.32	.94 ^b	2 min
Simon Squared	321	57.38	15.26	-1.14	2.00	.97 ^b	2 min
Antisaccade	306	0.79	0.13	-0.54	-0.60	.87 ^a	8 min
FlankerDL	316	729.95	299.27	1.86	2.89	.89 ^a	10 min
SACT	323	0.86	0.17	-1.50	1.32	.95 ^a	18 min
Selective Visual Arrays	291	1.60	1.06	0.57	-0.32	.58 ^b	---
Raven's Matrices	344	9.14	3.41	-0.36	-0.27	.81 ^a	---
Letter Sets	337	15.80	4.95	-0.17	-0.51	.89 ^a	---
Number Series	331	8.81	3.26	-0.15	-0.72	.83 ^a	---
Symmetry Span	327	26.41	11.58	0.01	-0.54	.81 ^a	---
Rotation Span	330	22.41	11.98	0.43	-0.08	.85 ^a	---
Mental Counters	316	76.40	14.12	-1.21	1.49	.91 ^a	---

Note. ^a Cronbach's alpha, ^b Split-half reliability with Spearman-Brown correction. Time = average administration time from starting to finishing the task. --- = administration time was not measured for this task.

Correlations

Task-level correlations are presented in Table 3. As can be seen, the three Squared tests correlated very highly with each other (average $r = .51$, correlations ranged from $r = .50$ to $r = .52$), demonstrating convergent validity. For comparison, the other four attention control tests (i.e., antisaccade, FlankerDL, SACT, and selective visual arrays) had numerically lower intercorrelations (average $r = .23$ after reversing the sign of FlankerDL). FlankerDL demonstrated weaker-than-expected correlations with most of the cognitive ability measures. After removing FlankerDL, the remaining other three attention control tests demonstrated better convergent validity (average $r = .38$). As expected, the three tests of fluid intelligence correlated significantly with each other (average $r = .55$), as did the tests of working memory capacity (average $r = .46$).

Table 3. *Task-level correlation matrix for Study 1.*

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Stroop Squared	---											
2. Flanker Squared	.50	---										
3. Simon Squared	.50	.52	---									
4. Antisaccade	.36	.40	.32	---								
5. FlankerDL	-.02	-.14	-.15	-.05	---							
6. SACT	.30	.33	.32	.41	-.13	---						
7. Selective Visual Arrays	.34	.32	.30	.34	-.06	.39	---					
8. Raven's Matrices	.44	.55	.33	.38	-.04	.47	.42	---				
9. Letter Sets	.36	.47	.36	.23	-.13	.36	.27	.54	---			
10. Number Series	.37	.43	.32	.24	-.14	.31	.33	.51	.60	---		
11. Symmetry Span	.16	.21	.18	.29	-.02	.20	.34	.28	.26	.21	---	
12. Rotation Span	.22	.28	.16	.22	-.02	.28	.33	.36	.32	.26	.63	---
13. Mental Counters	.32	.33	.30	.29	-.06	.44	.33	.42	.40	.37	.33	.43

Note. Boldface indicates $p < .05$. For these pairwise correlations, N ranges from 256 to 336 (listwise $N = 205$).

Exploratory Factor Analysis

We conducted an exploratory factor analysis to determine the latent structure underlying the ability measures (Table 4). We used principal axis factoring with an oblique promax rotation and pairwise deletion, and then extracted factors with eigenvalues greater than one. We extracted three factors that appeared to represent attention control, fluid intelligence, and working memory capacity. All of the attention control tasks except for FlankerDL had their highest loadings on the first factor and relatively low cross-loadings on the other two factors, providing further evidence for convergent validity of the three Squared tests. The second factor appeared to represent fluid intelligence and was primarily defined by letter sets and number series, and, to a lesser extent, Raven's matrices. The third factor appeared to represent working memory capacity and was primarily defined by symmetry span and rotation span, and, to a lesser extent, mental counters. Noteworthy cross-loadings included Raven's matrices loading on the first factor (i.e., attention control) and selective visual arrays loading on the third factor (i.e., working memory capacity).

The three factors were moderately correlated (Factor 1 with Factor 2: $r = .71$; Factor 1 with Factor 3: $r = .51$, Factor 2 with Factor 3: $r = .51$).

Table 4. *Exploratory factor analysis for Study 1.*

Measure	Factor 1	Factor 2	Factor 3
Stroop Squared	.66	.07	-.10
Flanker Squared	.59	.23	-.09
Simon Squared	.67	.05	-.14
Antisaccade	.67	-.21	.14
FlankerDL	-.05	-.17	.08
SACT	.41	.11	.14
Selective Visual Arrays	.41	-.02	.27
Raven's Matrices	.32	.41	.11
Letter Sets	-.11	.88	.02
Number Series	-.01	.75	-.02
Symmetry Span	.02	-.11	.78
Rotation Span	-.11	.04	.85
Mental Counters	.19	.22	.32
Eigenvalues	4.91	1.38	1.03

Note. Principal axis factoring with promax (oblique) rotation. Boldface indicates the strongest loading for each measure as well as any substantial cross-loadings.

Confirmatory Factor Analyses

Next, we conducted a series of confirmatory factor analyses using maximum likelihood estimation. We first created a model in which all the attention control measures were specified to load on a common factor. The model fit the data well ($\chi^2(14) = 14.70$, $p = .42$; CFI = .998, TLI = .997, RMSEA = .012, 90% CI [.000, .068], SRMR = .036) and is depicted in Figure 8. The three Squared tests had the highest loadings on the factor, ranging from .64 to .68. The other attention control measures had slightly lower loadings: antisaccade (.57), SACT (.46), selective visual arrays (.51). The exception was FlankerDL, which had a non-significant loading ($-.10$, $p = .22$) on the attention control factor. We elected to drop FlankerDL from subsequent models, as it did

not correlate significantly with most of the measures in the study, did not load on the common attention control factor, and contributed nothing to the questions being asked here.

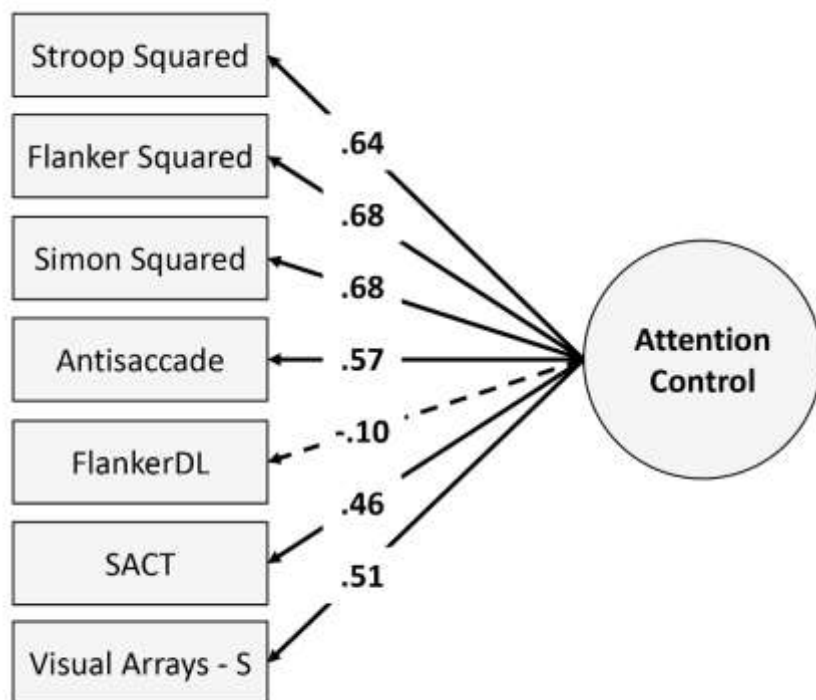


Figure 8. Latent variable model with all attention control measures loading on a common factor (Study 1).

To test how much variance the Squared tests and the other attention control tests shared at the latent level, we specified a model with two correlated factors, one for each group of attention tasks. The model is depicted in Figure 9 and fit the data well ($\chi^2(8) = 2.76, p = .949$; CFI = 1.00, TLI = 1.04, RMSEA = .000, 90% CI [.000, .006], SRMR = .017). We note two observations regarding this model. First, the factor loadings for the three Squared tests were slightly higher—ranging from .66 to .71—than the loadings for the other attention control tests, which ranged from .52 to .63. Second, the correlation between the two latent factors was .80, indicating that the two attention control factors shared 64% of their reliable variance, a statistically and practically significant amount that provides further evidence for the convergent

validity of the three Squared tests as measures of attention control. That said, setting the correlation between the latent factors equal to 1 resulted in significantly worse model fit, $\Delta\chi^2(1) = 9.13, p = .003$, raising the possibility that the three Squared tests and the three older tests captured some unique, potentially theoretically relevant variance.

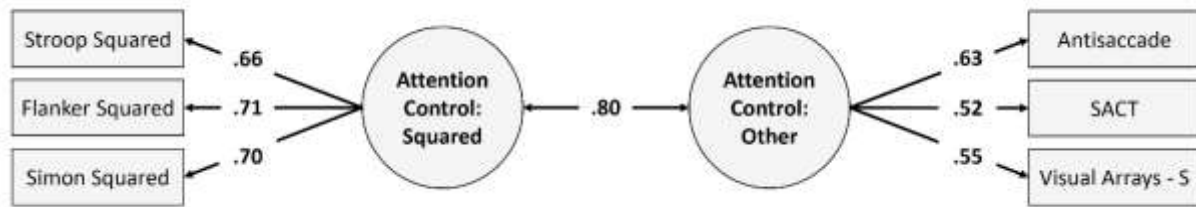


Figure 9. Latent variable model with the three Squared tests loading on one factor and the other attention control tests loading on another factor (Study 1).

In our next set of analyses, we examined correlations between attention control, working memory capacity, and fluid intelligence at the latent level. We created two attention control latent factors, one for the three Squared tests and one for the other tests of attention control. The purpose was to examine how correlations between attention control and the latent cognitive ability factors differed depending on how attention control was measured.

As shown in Figure 10, the Squared attention control factor correlated $r = .49$ with working memory capacity, whereas the other attention control factor correlated $r = .59$. The Squared attention control factor correlated $r = .81$ with fluid intelligence, whereas the other attention control factor correlated $r = .76$ with fluid intelligence. Fluid intelligence and working memory capacity correlated $r = .63$. The fit of the model was adequate ($\chi^2(48) = 120.88, p < .001$; CFI = .903, TLI = .866, RMSEA = .085, 90% CI [.066, .104], SRMR = .072).

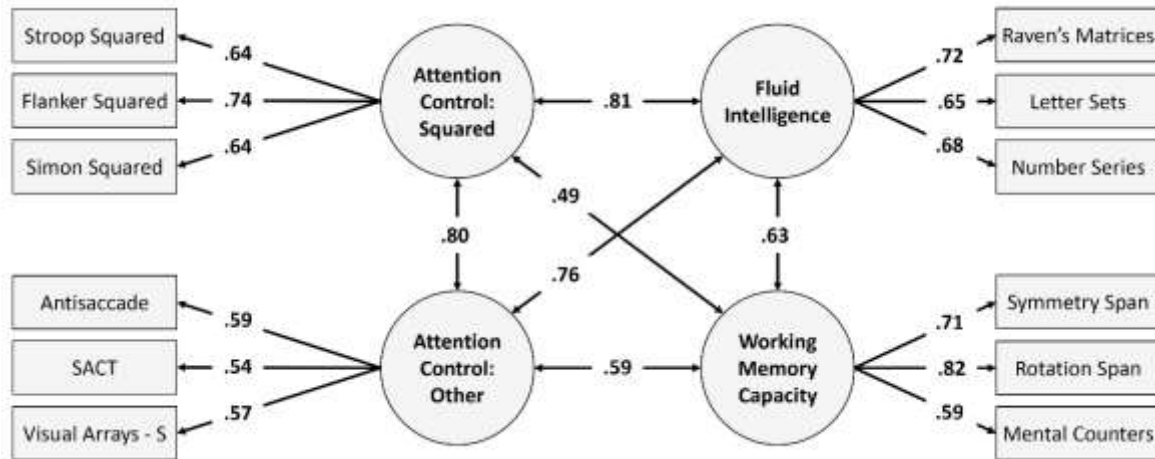


Figure 10. Correlated factors model with a Squared attention control factor and another attention control factor, each of which was allowed to covary with fluid intelligence and working memory capacity (Study 1).

Structural Equation Modeling

Next, we tested a series of structural equation models to determine the degree to which attention control—and the three Squared tests in particular—accounted for the covariance between working memory capacity and fluid intelligence. We tested two models, one in which attention control was identified using the three Squared tasks and another in which we used the other three attention control tests. In each model, attention control was specified as a predictor of fluid intelligence and working memory capacity, and the residuals of fluid intelligence and working memory capacity (representing the variance in each construct that remained after accounting for attention control) were allowed to correlate.

As shown in Figure 11, the three Squared tests were significant predictors of fluid intelligence ($\beta = .83, p < .001$) and working memory capacity ($\beta = .49, p < .001$) when modeled at the latent level. The correlation between the residuals of fluid intelligence and working memory capacity was significant, $r = .40, p < .001$. The model fit the data adequately ($\chi^2(24) = 79.16, p < .001$; CFI = .924, TLI = .887, RMSEA = .096, 90% CI [.073, .120], SRMR = .076).

We tested whether the residual correlation between fluid intelligence and working memory capacity after accounting for attention control was significantly weaker than the latent bivariate correlation between these factors ($r = .40$ vs. $r = .63$, see Figure 11). Setting the residual correlation equal to $r = .63$ significantly worsened model fit, $\Delta\chi(1) = 4.235$, $p = .040$, indicating that the Squared attention control factor accounted for a significant proportion of the covariance between fluid intelligence and working memory capacity.

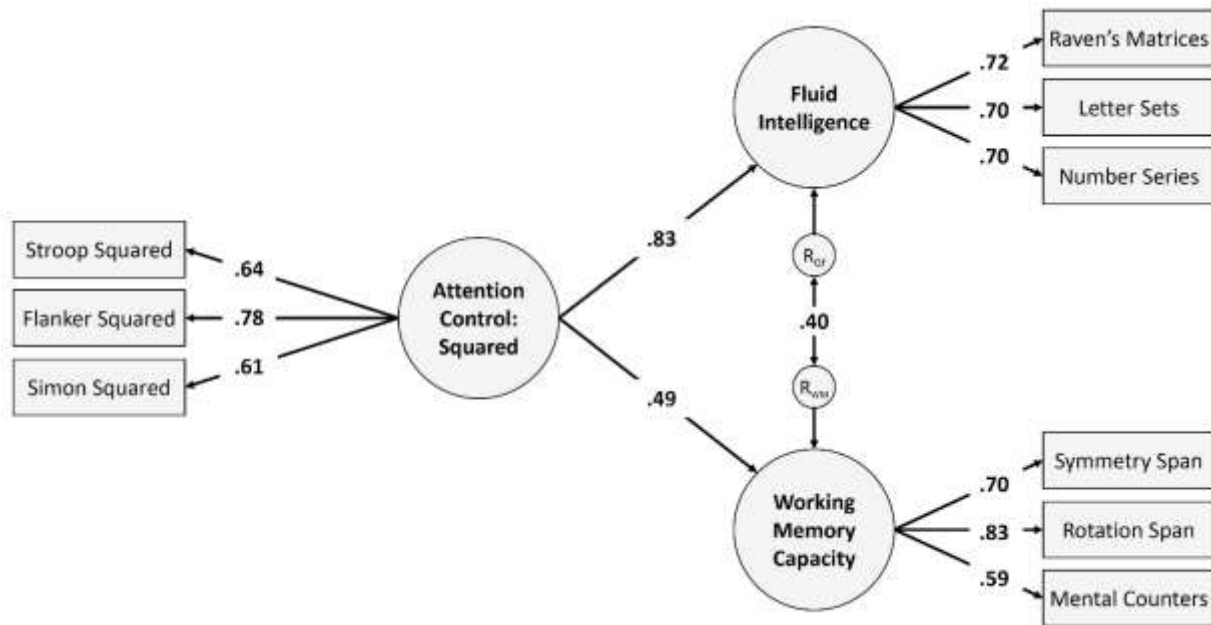


Figure 11. Structural equation model with a latent factor for the three Squared tests of attention control predicting fluid intelligence and working memory capacity. R_{Gf} and R_{WM} represent the residual variance in fluid intelligence and working memory capacity, respectively, after accounting for attention control. The correlation between the residuals R_{Gf} and R_{WM} was significant, $r = .40$, $p < .001$, but significantly weaker than the correlation between fluid intelligence and working memory capacity before accounting for attention control ($r = .63$), $\Delta\chi(1) = 4.235$, $p = .040$ (Study 1).

As shown in Figure 12, the other attention control tests were significant predictors of fluid intelligence ($\beta = .72$, $p < .001$) and working memory capacity ($\beta = .57$, $p < .001$) when modeled at the latent level. The correlation between the residuals of fluid intelligence and

working memory capacity was significant, $r = .24$, $p = .042$. The model fit the data adequately ($\chi^2(24) = 77.13$, $p < .001$; CFI = .910, TLI = .865, RMSEA = .095, 90% CI [.071, .119], SRMR = .073). We tested whether the residual correlation between fluid intelligence and working memory capacity after accounting for attention control was significantly weaker than the latent bivariate correlation between these variables ($r = .24$ vs. $r = .63$, see Figure 11). Setting the residual correlation equal to $r = .63$ significantly worsened model fit, $\Delta\chi(1) = 14.25$, $p < .001$, indicating that the other attention control tests, when modeled at the latent level, accounted for a significant proportion of the covariance between fluid intelligence and working memory capacity.

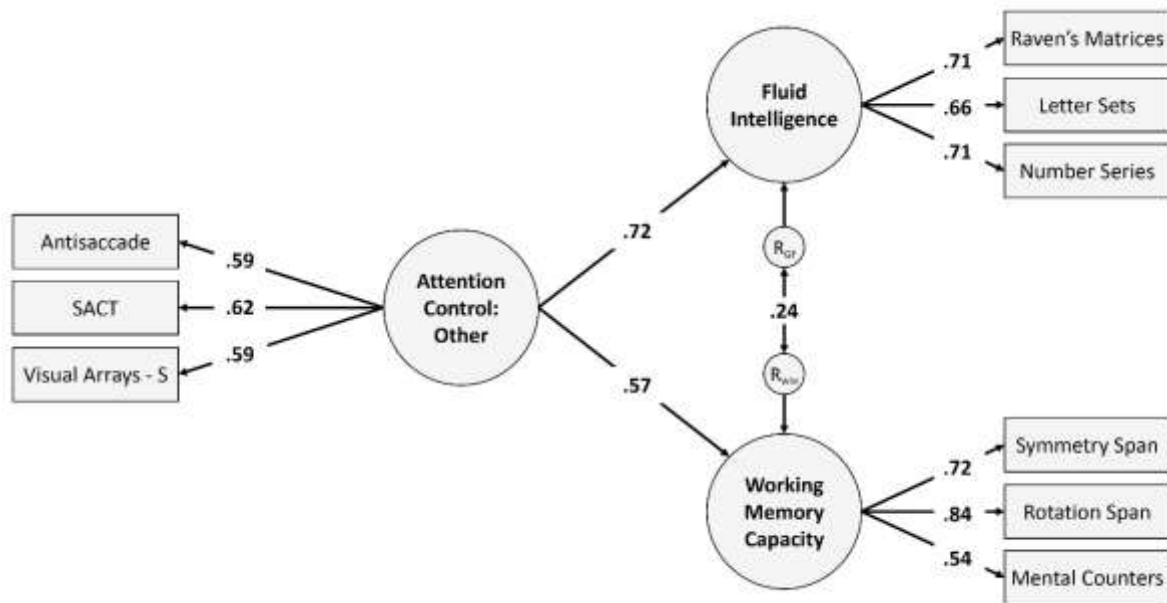


Figure 12. Structural equation model with a latent factor for the three non-Squared tests of attention control predicting fluid intelligence and working memory capacity. R_{GF} and R_{WM} represent the residual variance in fluid intelligence and working memory capacity, respectively, after accounting for attention control. The correlation between the residuals R_{GF} and R_{WM} was significant, $r = .24$, $p = .042$, but significantly weaker than the correlation between fluid intelligence and working memory capacity before accounting for attention control ($r = .63$), $\Delta\chi(1) = 14.25$, $p < .001$ (Study 1).

Analysis of Trial Types

As we report in the Supplemental Materials, we found that participants earned the most points and responded the quickest on Fully Congruent trials. Repeated measures ANOVAs indicated that there were significant main effects and interaction effects between stimulus congruency, response option congruency, and their interaction, on points earned and response times for each Squared task except for Simon Squared. In Simon Squared, only the interaction effects were significant. However, there was no other consistent pattern of differences in performance on the trial types of the three Squared tasks, and no consistent pattern of differences in correlations between performance on the trial types and other cognitive abilities. See Supplemental Materials for analyses.

Discussion of Study 1

Study 1 established that the three-minute “Squared” tests of attention control—Stroop Squared, Flanker Squared, and Simon Squared—demonstrate strong psychometric properties. Specifically, we found compelling evidence for the internal consistency reliability of all three tasks, with split-half reliabilities ranging from .93 to .97. We also found strong evidence for the Squared tasks’ construct validity, with patterns of correlations indicating that the Squared attention control tests correlated very highly with each other at the observed level (average $r = .51$), and very highly with the best attention control tests to emerge from Draheim et al. (2021) at the latent level ($r = .80$, after dropping FlankerDL due to a non-significant loading on the attention control factor). Finally, we found that the three Squared tests of attention control can be used to predict individual differences in complex cognition at the latent level, with a large predictive path to fluid intelligence ($\beta = .83$; $R^2 = 69\%$) and a moderate predictive path to working memory capacity ($\beta = .49$; $R^2 = 24\%$).

When testing whether attention control accounts for the positive covariation between fluid intelligence and working memory capacity, we found that regardless of how we specified the latent attention control factor, it significantly reduced the correlation between fluid intelligence and working memory capacity, but did not completely eliminate it (residual correlations ranged from $r = .24$ to $r = .40$ depending on how the latent attention control factor was defined). We note that this pattern of results is not entirely surprising, because [Draheim et al. \(2021\)](#) also found that attention control rarely fully accounted for the positive correlation between working memory capacity and fluid intelligence; in most of the models that tested different combinations of attention control indicators, the correlation between working memory capacity and fluid intelligence remained statistically significant.

As we noted in the Introduction, [Martin et al. \(2021\)](#) found that selective visual arrays often cross-loads on attention control and working memory capacity factors. We found the same pattern of results in an exploratory factor analysis (see Table 5). We also found suggestive evidence that the Squared tasks may account for slightly less of the covariance in the fluid intelligence-working memory capacity relationship than the other three attention control tests (compare the residual correlations of $r = .40$ vs. $r = .24$). It is possible that the short-term storage demands of selective visual arrays increased the latent correlation between the attention control factor and working memory capacity, allowing it to account for more of the variance that was shared between working memory capacity and fluid intelligence. That said, when we set the residual correlation between fluid intelligence and working memory capacity to .24 after accounting for attention control (using the Squared tasks as indicators), the reduction in model fit was not statistically significant, $\Delta\chi(1) = 1.718, p = .19$. Thus, the other attention control tests did

not account for significantly more of the covariation between fluid intelligence and working memory capacity than the Squared tasks did when modeled at the latent level.

When conducting analyses on the different trial types in the Squared tasks, we found that participants earned more points and responded quickest on fully congruent trials, suggesting that these trials were easier for participants. There was no pattern of performance differences across the other trial types that was consistent across all three Squared tasks. When examining correlations between performance on each trial type and cognitive abilities, we found relatively inconclusive evidence that correlations diverged based on the degree of cognitive control required by each trial type (see Supplemental Materials).

Study 2: In-Laboratory Study

There are a few limitations of Study 1 that motivated the need for Study 2. For example, Study 1 used an online sample of participants recruited through Prolific and MTurk. Although we employed data filtering procedures to eliminate problematic data points and outliers, Study 2 circumvented concerns about the validity of online data collection by recruiting a large sample of participants from Georgia Tech and the surrounding Atlanta community, and testing them in our laboratory under the supervision and guidance of trained research assistants. Another limitation of Study 1 is that while we included several measures of different cognitive abilities, we did not measure participants' processing speed or performance on more complex multitasks such as those that are used as a proxy for real-world work (see Martin et al., 2020; Burgoyne, Hambrick, & Altmann, 2021). Study 2 addressed these limitations by including multiple measures of both processing speed and multitasking. Throughout Study 2 we note occasions when the results are broadly consistent (or inconsistent) with the results of Study 1.

Method

Participants

The study was conducted at the Georgia Institute of Technology in Atlanta, GA, USA. All participants were required to be native English speakers and 18-35 years of age. No screening on vision or hearing was performed. We recruited participants from Georgia Tech, other surrounding colleges in Atlanta, and the broader Atlanta community. Georgia Tech students enrolled in an undergraduate psychology course were given the option to receive 2.5 hours of course credit or monetary compensation for each session. This study was approved by the Georgia Institute of Technology's Institutional Review Board under Protocol H20165. A total of 327 subjects completed at least four sessions. Therefore, our sample should be large enough for stable estimates of correlations (i.e., $N > 250$) (Schönbrodt & Perugini, 2013).

Procedure

Data were collected as part of a larger project, which consisted of more than 40 cognitive tasks administered over five sessions lasting 2.5 hours each. We included participants who completed the first four sessions of the study, because the fifth session consisted of tasks not relevant to the present work. We report on a subset of the data, focusing specifically on the same tasks that were used during Study 1 (i.e., the online study), as well as tests of processing speed and multitasking paradigms that serve as criterion measures. Further information regarding the scope of the data collection effort and other research products based on it can be found at the following link: <https://osf.io/qbwem>.

Participants scheduled each study session according to their own availability, but they were not allowed to complete more than one session on a given day. Participants were paid \$200 for completing the five in-laboratory sessions (\$30 for session 1, \$35 for session 2, \$40 for session 3, \$45 for Session 4, and \$50 for Session 5). We additionally offered participants who

completed Session 5 the opportunity to complete an online follow-up study, which included the same tasks as in Study 1, for \$50. Georgia Tech students were allowed to choose a combination of either financial compensation or research participation credits—the latter is required by some undergraduate psychology courses at Georgia Tech. Participants who frequently rescheduled, missed appointments, or regularly failed to follow directions were not invited back for subsequent sessions.

During data collection, participants were seated in individual testing rooms with a research assistant assigned to proctor each session. The research assistant's job was to run each cognitive test, ensure the participant understood the instructions, and make sure participants were following the rules of the lab, such as not using their phone during the study. The research assistants took extensive notes on participant conduct, which was used to make decisions about data exclusions described below. Up to 7 participants could be tested in a given session, although typically 2-4 participants were scheduled for each timeslot.

Online Follow-Up Study. Participants who completed the in-lab study were offered the opportunity to complete additional computerized tasks—the same as those used in Study 1—using their personal computers outside of the laboratory. The purpose of this data collection effort was to collect test-retest-*retest* reliability data on each of the three Squared tasks across different testing environments, and using E-Prime Go. To be clear, in the in-lab version of the study, participants completed the three Squared tasks twice; once during Session 1 and once during Session 4, separated by an average of 32 days. Thus, by conducting this online follow-up study, separated by an average of 65 days from the second in-lab test, we obtained a third measure of performance on the Squared tasks, this time in a different testing environment.

Demographics

Participants were asked to report their age, gender, and ethnicity. They were asked whether English was the first language they learned and the age at which they learned it, and whether they were fluent in other languages. Participants were asked to report the highest level of education they had achieved as well as their annual household income. Participants were asked whether they had corrected vision, and also whether they had any conditions (e.g., illness, disability, medication use) that might affect their performance on cognitive tasks.

Attention Control

Stroop Squared. See Study 1. Participants completed Stroop Squared up to three times over the course of the study: once during Session 1, once during Session 4, and once during the online follow-up study which occurred after all five in-lab sessions were completed.

Flanker Squared. See Study 1. Participants completed Flanker Squared up to three times over the course of the study: once during Session 1, once during Session 4, and once during the online follow-up study which occurred after all five in-lab sessions were completed.

Simon Squared. See Study 1. Participants completed Simon Squared up to three times over the course of the study: once during Session 1, once during Session 4, and once during the online follow-up study which occurred after all five in-lab sessions were completed.

Antisaccade. See Study 1.

Flanker Adaptive Deadline (FlankerDL). See Study 1.

Sustained Attention to Cue (SACT). See Study 1.

Selective Visual Arrays. See Study 1.

Fluid Intelligence

Raven's Advanced Progressive Matrices. See Study 1.

Letter Sets. See Study 1.

Number Series. See Study 1.

Working Memory Capacity

Advanced Symmetry Span. See Study 1.

Advanced Rotation Span. See Study 1.

Mental Counters. See Study 1.

Processing Speed

Digit String Comparison (Redick et al., 2012). Participants viewed strings of three, six, or nine digits appearing to the left and right of a central line. The digit strings could either be the same or differ by a single digit. Participants indicated their response by clicking on a button on the screen labeled SAME for identical strings or DIFF for mismatching strings. After completing six practice trials, participants completed as many trials as possible within two 30s blocks of the task. The measure of performance was the number of accurate responses across both blocks, minus the number of inaccurate responses.

Letter String Comparison (Redick et al., 2012; Salthouse & Babcock, 1991). The Letter String Comparison task was identical to the Digit String Comparison task, except that participants viewed and made judgements about strings containing three, six, or nine characters.

Pattern Comparison (Salthouse & Babcock, 1991). Participants were presented with two patterns (i.e., symbols made from simple line drawings) on either side of a line and used the keyboard to indicate whether they were the same or different. Participants completed as many trials as possible within two 30s blocks of the task. The measure of performance was the number of accurate responses across both blocks, minus the number of inaccurate responses.

Multitasking Paradigms

Synthetic Work for Windows (SynWin; Elsmore, 1994; Figure 13). Participants interleaved four subtasks. In the memory search subtask, a list of six letters was presented and then disappeared. A series of probe letters was then displayed. Participants determined whether each probe letter came from the original list or not; they were awarded 10 points for correct responses and penalized 10 points for incorrect responses or failures to respond. In the math subtask, participants added two three-digit numbers and clicked a “Done” button to register their answer. Participants were awarded 20 points for correct responses and penalized 10 points for incorrect responses. They had 20 seconds to answer each problem; every second after 20 seconds, 10 points were subtracted from their total score. In the visual monitoring subtask, participants clicked a “fuel gauge” before the indicator on the gauge reached zero; they earned more points for clicking the fuel gauge when the indicator was near zero, but lost 10 points for every second that the indicator was at zero. In the auditory monitoring subtask, participants responded to a high-pitched tone and ignored a low-pitched tone; they were awarded 10 points for correct responses and penalized 10 points for incorrect responses or failures to respond. Participants were presented task instructions and then completed three 5-minute test blocks of SynWin. Scores from the three test blocks were averaged to create an overall score.

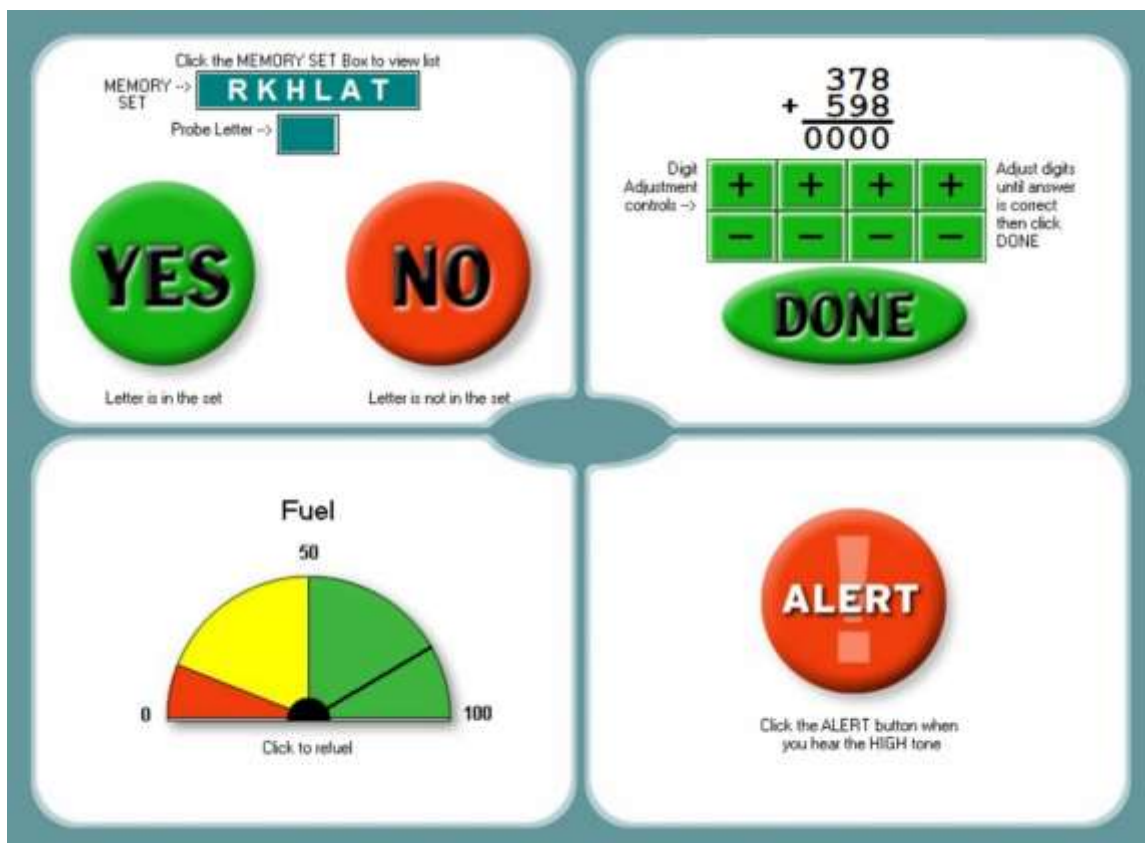


Figure 13. Synthetic Work (SynWin). The four subtasks are: Memory Search (top-left); Math (top-right); Visual Monitoring (bottom-left); and Auditory Monitoring (bottom-right).

Foster Multitask (Martin et al., 2020; Figure 14). The Foster Task is a multitask modeled after SynWin. The dependent variable was the aggregate score across all four subtasks. Participants increased their score by responding appropriately within a brief period, resulting in 100 points being added to the score for that subtask. Errors led to a subtraction of 100 points, while failing to respond within a brief response window resulted in a drop in points for that task until a response was made or the counter reached zero. Participants completed three five minute blocks of the task. The subtasks were: 1) Telling time: Participants saw an unlabeled analog clock face and were asked to choose, via mouse click, which of four alternatives corresponded to the position of the clock's hour and minute hand. 2) Visual monitoring: Participants were

presented with a disk that would start spinning periodically. They were instructed to click on the disk as soon as possible to make it stop moving. 3) Word recall: Participants were prompted to remember a word printed in green. After a brief display, the word was removed. Shortly thereafter, a new prompt would appear in red asking whether this new word was the one they were supposed to remember previously. They responded either yes or no by mouse click and a new green to-be-remembered word was presented. 4) Math problems: Participants solved simple two term math problems. They were either addition, subtraction, multiplication, or division, and the participant chose the correct answer from among four possible answers.



Figure 14. A labeled snapshot of the Foster Multitask interface.

Control Tower (Redick et al., 2016; Figure 15). The control tower task was designed with one primary task that was completed while also managing interruptions from four other tasks. Both visual and auditory processing was necessary for task success. Participants were presented with a visual display with various subtasks to complete. For the primary task, participants searched through an array of numbers, letters, and symbols on the left side of the screen and selected the appropriate items from the array on the right side of the screen, as

indicated by task instructions: for numbers, participants clicked on the identical numbers in the right array; for letters, participants clicked on the letter that preceded it alphabetically in the right array; and for symbols, participants referred to a consistently mapped symbol *code book* and clicked on the relevant symbols in the right array.

Participants were instructed to complete as many array comparisons as possible over the duration of the task. Meanwhile, distractor tasks (radar, airplane, color, problem-solving) were presented visually and via headphones to reference items listed below the arrays at specific times or when specific colors were presented. For the radar task, participants were instructed to click either the *inside* or *outside* button below the radar when a blip occurred. For the airplane task, requests for landing on one of three runways were presented via headphones and the participant either agreed to or denied the request, according to the availability listed in the button marked *runway*. For the color task, a color flashed briefly and the participant pressed one of three “error” buttons according to the mapping presented in the button marked *protocol*. For the problem-solving task, trivia, logic, and general knowledge questions were presented via headphones, and participants clicked on one of three possible answers presented at the bottom of the screen. Participants completed one 10-min block of the task. The participant’s primary score was the number of correct number, letter, and symbol comparisons completed (no theoretical maximum score). The participant’s distractor score was the sum of the correct decisions across the various

NATURE AND MEASUREMENT OF ATTENTION CONTROL

distractor subtasks (out of 30). Both the primary and distractor scores were used as outcome measures.



Figure 15. Labeled snapshot of the Control Tower interface.

Data Preparation

We used the same data preparation procedure as in Study 1. That is, we removed participants' scores on a task if they showed severely poor performance indicating they did not understand the instructions or were not performing the task as intended. Specifically, we computed chance-level performance on each task; any scores that were at or below chance-level performance were identified as problematic data points and set to missing. This procedure was applied to the three Squared tests of attention control, antisaccade, selective visual arrays, SACT, and FlankerDL. We did not remove problematic data points for the three tests of fluid intelligence. For FlankerDL, we set trial-level performance to missing if the response time on that trial was less than 200ms, on the basis that these responses were too fast and likely

represented mis-clicks. For the Advanced Span tasks, problematic data points were defined by chance-level performance or worse on the processing subtask. After removing problematic data points from the in-lab sample (28) and the online follow-up sample (29), we performed a two-pass outlier exclusion procedure. We removed data points that were more than 3.5 standard deviations worse than the sample mean two times, recomputing the sample mean and standard deviation each time. On the first pass, the outlier exclusion process removed 24 data points from the in-lab sample and 1 data point from the online follow-up sample. On the second pass, the outlier exclusion process removed 15 data points from the in-lab sample and 0 data points from the online follow-up sample.

Modeling Approach and Fit Statistics

We used the same modeling approach and fit statistics as in Study 1.

Results

Demographic information is reported in Table 5. The participants' average age was 22 ($SD = 4$) years old and a majority were female (58.9%). Our in-lab sample was slightly older than the online sample from Study 1, which had a mean age of 27—the difference was statistically significant ($t(660) = 14.80, p < .001$). In terms of race/ethnicity, 41% of the sample identified as Asian or Pacific Islander, 28% identified as White, 13% identified as Black or African American, and the remainder selected “Other” or declined to respond. The majority of participants (90.8%) had attended at least some college.

Table 5. *Demographic Information for Study 2*

Demographic	Statistic
Age (years)	Mean = 21.95 $SD = 4.09$ Range = 18 – 35

NATURE AND MEASUREMENT OF ATTENTION CONTROL

Gender	Male = 39.5% Female = 58.9% Self-Identify/Other = 1.3% Transgender Male = 0.3%
At least some college?	Yes: 90.8% No: 9.2%
Ethnicity	White: 28.3% Black or African American: 13.4% Asian or Pacific Islander: 41.4% Other* : 16.9%

Note. $N = 314$. *Other includes, Hispanic or Latino, Native American, and Other.

Descriptive statistics are reported in Table 6. Estimates of internal consistency reliability were very high for all administrations of the Squared tasks, ranging from .94 to .97. The other tasks also had adequate to excellent internal consistency reliability, ranging from .73 to .95. Data transformations were not performed on the outcome measures because skewness and kurtosis were acceptable for all measures.

Table 6. *Descriptive statistics for Study 2.*

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Skew</i>	<i>Kurtosis</i>	<i>Reliability</i>	<i>Time</i>
Stroop Squared 1	311	41.44	12.76	-0.11	0.17	.94 ^b	2 min
Stroop Squared 2	311	39.95	13.18	-0.29	0.52	.94 ^b	2 min
Stroop Squared 3	63	41.06	16.76	-0.98	0.48	.95 ^b	2 min
Flanker Squared 1	290	40.43	14.17	0.10	-0.20	.97 ^b	2 min
Flanker Squared 2	290	39.47	14.13	0.14	0.26	.96 ^b	2 min
Flanker Squared 3	63	31.83	15.80	-0.15	-0.59	.94 ^b	2 min
Simon Squared 1	310	67.87	9.34	-0.20	0.24	.94 ^b	2 min
Simon Squared 2	310	67.32	9.03	-0.22	-0.07	.95 ^b	2 min
Simon Squared 3	61	63.85	12.91	-0.85	1.31	.96 ^b	2 min
Antisaccade	299	0.81	0.12	-0.62	-0.64	.87 ^a	---
FlankerDL	307	660.80	273.14	1.79	2.72	.89 ^a	9 min
SACT	307	0.89	0.10	-1.11	0.71	.87 ^a	17 min
Visual Arrays	316	2.47	0.70	-0.51	0.07	.91 ^b	12 min
Raven's Matrices	316	11.30	2.87	-0.41	-0.26	.77 ^a	---
Letter Sets	312	16.41	4.41	-0.17	-0.69	.85 ^a	---
Number Series	317	9.99	2.98	-0.22	-0.73	.73 ^a	---

NATURE AND MEASUREMENT OF ATTENTION CONTROL

Symmetry Span	310	29.90	9.73	-0.24	-0.40	.76 ^a	---
Rotation Span	310	25.16	8.66	-0.11	-0.20	.73 ^a	---
Mental Counters	305	79.26	13.63	-1.24	1.13	.91 ^a	---
Digit Comparison	307	29.90	5.51	-0.45	0.02	.88 ^b	---
Letter Comparison	307	20.53	4.10	0.12	0.39	.82 ^b	---
Pattern Comparison	306	39.06	6.01	-0.09	-0.22	.94 ^b	---
SynWin	308	3243.83	568.57	-0.53	1.14	.90 ^a	---
Foster Multitask	302	96011.66	26343.41	-0.20	0.05	.95 ^a	---
Control Tower (P)	306	102.55	30.62	-0.02	0.34	---	---
Control Tower (D)	306	25.83	2.45	-0.95	0.46	---	---

Note. ^a Cronbach's alpha, ^b Split-half reliability with Spearman-Brown correction, --- = internal consistency reliability could not be computed for Control Tower; administration time was not measured for these tasks. Time = average administration time from starting to finishing the task. The number following each "Squared" task name indicates the test administration number. Control Tower (P) = Primary score; (D) = Distractor score.

Next, we computed test-retest-retest reliability for the Squared tasks by correlating performance on the first attempt (during Session 1) with performance on the second attempt (during Session 4; on average 32 days [$SD = 33$] after the first administration) and third attempt (during the online follow-up study; approximately 65 days [$SD = 57$] after the second administration). The test-retest reliabilities ranged from good to excellent, as shown in Table 7. Specifically, examining performance on the first and second administrations of the test revealed very high correlations: Stroop Squared ($r = .53, p < .001$), Flanker Squared ($r = .74, p < .001$), Simon Squared ($r = .75, p < .001$). We observed a similar pattern comparing performance on the second and third administrations, despite the fact that the third administration of the test occurred outside of the laboratory on the participants' personal computers: Stroop Squared ($r = .55, p < .001$), Flanker Squared ($r = .46, p < .001$), Simon Squared ($r = .49, p < .001$). Thus, these results indicate that the three Squared tasks have good to very-high test-retest-retest reliability, even across testing environments. In other words, participants who performed well during the first

administration of the test performed well in latter administrations of the test, and participants who performed poorly tended to continue performing poorly.

Table 7. *Correlations between each administrations of the Squared tasks from Study 2.*

Measure	Stroop ² 1	Stroop ² 2	Stroop ² 3	Flanker ² 1	Flanker ² 2	Flanker ² 3	Simon ² 1	Simon ² 2	Simon ² 3
Stroop ² 1	---								
Stroop ² 2	.53	---							
Stroop ² 3	.49	.55	---						
Flanker ² 1	.48	.39	.50	---					
Flanker ² 2	.38	.49	.34	.74	---				
Flanker ² 3	.24	.29	.53	.45	.46	---			
Simon ² 1	.51	.28	.33	.53	.40	.35	---		
Simon ² 2	.41	.48	.32	.46	.50	.41	.75	---	
Simon ² 3	.30	.43	.60	.44	.28	.69	.48	.49	---

Note: The ² symbol is used as an abbreviation for the task name (i.e., Stroop² = Stroop Squared). The number following each task name indicates the test administration number. Bold = Statistically significant ($p < .05$). N ranges from 287 to 311 for everything except correlations involving the third administration of each of the Squared tasks (N s for those tasks ranged from 55 to 59).

Next, we tested whether there were practice effects on the Squared tasks by examining within-subject changes in performance across testing administrations. As shown in Figure 16, participants performed about the same on the task each time they completed it. Comparing the first attempt to the second attempt revealed very small differences in performance: Stroop Squared ($d = -0.11$, $t(310) = 2.09$, $p = .037$), Flanker Squared ($d = -0.07$, $t(289) = 1.60$, $p = .112$), Simon Squared ($d = -.06$, $t(309) = 1.49$, $p = .138$); negative values indicate that participants earned lower scores on the second administration of the test. Comparing the second attempt to the third attempt similarly revealed small differences in performance: Stroop Squared ($d = 0.22$, $t(56) = 1.76$, $p = .084$), Flanker Squared ($d = -.35$, $t(54) = 2.47$, $p = .017$), Simon Squared ($d = -.25$, $t(55) = 1.83$, $p = .072$). Thus, the Squared tasks demonstrated surprising resistance to

practice effects; within-participant changes in performance were very small and generally non-significant across testing administrations.

We conducted a repeated-measures ANOVA on each of the three Squared tasks to determine whether there was a main effect of repeated testing on overall performance. We note that these analyses are underpowered, as only a small subset of participants completed the task three times. The effect of test administration on Stroop Squared was non-significant, $F(2, 55) = 1.54$, $p = .223$, $\eta_p^2 = .053$; the effect of test administration on Flanker Squared was marginally significant, such that scores decreased over time, $F(2, 53) = 3.09$, $p = .054$, $\eta_p^2 = .104$; the effect of test administration on Simon Squared was marginally significant, such that scores decreased over time, $F(2, 54) = 3.16$, $p = .050$, $\eta_p^2 = .105$.

NATURE AND MEASUREMENT OF ATTENTION CONTROL

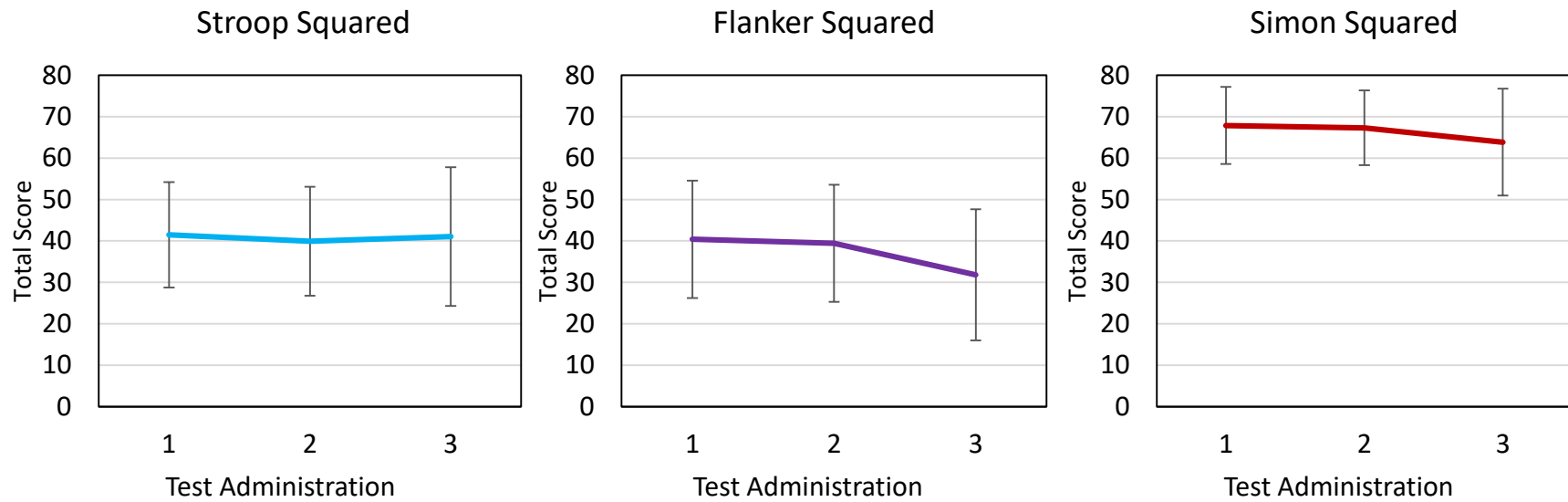


Figure 16. Scores on each of the three Squared tasks across the three test administrations. Error bars represent ± 1 SD around the mean (Study 2).

Task-Level Correlations

Task-level correlations are presented in Table 8. As was the case in Study 1, performance on the first attempt of each of the three Squared tests correlated very highly with each other (average $r = .50$, correlations ranged from $r = .48$ to $r = .53$), demonstrating convergent validity. For comparison, the other four attention control tests (i.e., antisaccade, FlankerDL, SACT, and selective visual arrays) had much lower intercorrelations (average $r = .22$ after reversing the sign of FlankerDL). As in Study 1, FlankerDL demonstrated near-zero correlations with most of the cognitive ability measures, and as a consequence we dropped FlankerDL from all subsequent analyses. After removing FlankerDL, the remaining three attention control tests demonstrated better convergent validity (average $r = .34$). The correlation between performance on the first attempt of each of the Squared tasks and the other three attention control tasks (i.e., antisaccade, SACT, and selective visual arrays) was $r = .33$, providing more evidence for the convergent validity of the Squared tasks. The three tests of fluid intelligence correlated significantly with each other (average $r = .48$), as did the tests of working memory capacity (average $r = .42$), tests of processing speed (average $r = .50$), and tests of multitasking ability (average $r = .44$).

Turning next to predictive validity at the bivariate level, the three squared tasks showed substantial and significant correlations with almost all of the other cognitive ability measures. Successive administrations of the Squared tasks did not appear to change their predictive validity much, which is consistent with our finding of high test-retest-retest reliability and limited practice effects. Specifically, the average correlation between all the non-Squared cognitive ability measures (except FlankerDL) and Stroop Squared 1 was $r = .34$; with Stroop Squared 2, the average correlation was $r = .35$; and with Stroop Squared 3, the average correlation was $r = .41$. The average correlation between the non-Squared cognitive ability measures (except

FlankerDL) and Flanker Squared 1 was $r = .40$; with Flanker Squared 2, the average correlation was $r = .37$; and with Flanker Squared 3, the average correlation was $r = .35$. The average correlation between the non-Squared cognitive ability measure and Simon Squared 1 was $r = .38$; with Simon Squared 2, the average correlation was $r = .40$; and with Simon Squared 3, the average correlation was $r = .33$. Thus, the three Squared tasks showed strong relationships with many of the cognitive ability measures at the observed level, and repeated testing on the tasks did little to compromise these relationships. In the next sections, we examine these relationships further at the construct-level by using a factor-analytic approach.

NATURE AND MEASUREMENT OF ATTENTION CONTROL

Table 8. Task-level correlation matrix for Study 2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1. Stroop Squared 1	---																								
2. Stroop Squared 2	.53	---																							
3. Stroop Squared 3	.49	.55	---																						
4. Flanker Squared 1	.48	.39	.50	---																					
5. Flanker Squared 2	.38	.49	.34	.74	---																				
6. Flanker Squared 3	.24	.29	.53	.45	.46	---																			
7. Simon Squared 1	.51	.28	.33	.53	.40	.35	---																		
8. Simon Squared 2	.41	.48	.32	.46	.50	.41	.75	---																	
9. Simon Squared 3	.30	.43	.60	.44	.28	.69	.48	.49	---																
10. Antisaccade	.35	.38	.21	.44	.41	.30	.32	.33	.18	---															
11. FlankerDL	.05	-.09	-.08	-.07	-.06	-.24	.07	.03	-.19	-.08	---														
12. SACT	.11	.23	.12	.26	.24	.34	.21	.26	.28	.30	-.08	---													
13. Visual Arrays	.38	.39	.58	.49	.41	.43	.38	.36	.54	.39	-.14	.32	---												
14. Raven's Matrices	.32	.36	.50	.43	.44	.36	.24	.24	.23	.29	-.12	.11	.44	---											
15. Letter Sets	.35	.34	.28	.42	.31	.44	.31	.36	.45	.26	-.01	.09	.32	.40	---										
16. Number Series	.38	.39	.50	.48	.44	.38	.39	.43	.45	.27	.00	.06	.44	.45	.58	---									
17. Symmetry Span	.28	.29	.42	.32	.31	.32	.28	.27	.13	.23	.01	.14	.43	.33	.30	.31	---								
18. Rotation Span	.27	.26	.48	.29	.32	.36	.25	.21	.23	.27	-.05	.14	.33	.29	.19	.29	.51	---							
19. Mental Counters	.36	.37	.58	.49	.49	.34	.39	.40	.32	.44	.00	.27	.54	.48	.42	.50	.39	.37	---						
20. Digit Comp.	.35	.32	.37	.37	.33	.18	.54	.56	.24	.31	-.06	.19	.34	.28	.44	.40	.24	.24	.34	---					
21. Letter Comp.	.29	.30	.22	.27	.33	.21	.41	.48	.09	.21	.06	.16	.26	.19	.43	.30	.23	.19	.31	.61	---				
22. Pattern Comp.	.41	.36	.61	.40	.37	.28	.49	.49	.47	.30	.11	.21	.43	.39	.32	.37	.33	.32	.38	.49	.41	---			
23. SynWin	.42	.44	.50	.49	.45	.45	.47	.51	.45	.37	-.02	.27	.43	.42	.54	.58	.37	.33	.46	.52	.40	.47	---		
24. Foster Multitask	.42	.43	.47	.54	.44	.48	.65	.66	.65	.38	-.01	.26	.48	.39	.52	.63	.34	.28	.47	.58	.47	.50	.62	---	
25. Control Tower (P)	.39	.43	.35	.35	.34	.44	.51	.55	.39	.30	.00	.17	.38	.33	.51	.54	.28	.21	.42	.52	.47	.46	.51	.61	---
26. Control Tower (D)	.33	.29	.30	.32	.34	.33	.22	.21	.24	.26	-.12	.10	.31	.31	.32	.35	.13	.15	.26	.22	.17	.32	.35	.34	.23

Note. Boldface indicates $p < .05$. For these pairwise correlations, N ranges from 273 to 312 (listwise $N = 200$) for everything except correlations involving the third administration of each of the Squared tasks (N s for those tasks ranged from 53 to 60). Control Tower (P) = Primary score; (D) = Distractor score.

Exploratory Factor Analysis

In the following sections, we use the participants' performance on the first test administration of each of the Squared tasks. We reasoned that participants did not receive practice on all the other tasks, so using participants' first attempt on the Squared tasks would make factor loadings more interpretable. We conducted an exploratory factor analysis to determine the latent structure underlying the ability measures (Table 9). We did not include the multitasking measures in this analysis because multitasking is a complex activity that likely demands multiple cognitive abilities, and perhaps as a result, its inclusion led to a factor structure that was very difficult to interpret. We used principal axis factoring with an oblique promax rotation to allow factors to correlate and extracted factors with eigenvalues greater than one.

We extracted four factors. They appeared to represent fluid intelligence (Factor 1), processing speed (Factor 2), attention control (Factor 3), and working memory capacity (Factor 4). The factors were highly correlated (r s ranged from .44 to .71; Table 10), which could explain why many measures had substantial cross-loadings on more than one factor. One particular issue that this exploratory factor analysis revealed is that mental counters had a low loading (.14) on the working memory capacity factor, and higher loadings on the fluid intelligence (.36) and attention control factors (.35). It appears that mental counters, which demands updating in a manner similar to a running span task, did not cohere with the two complex span measures of working memory capacity, symmetry span and rotation span. Thus, throughout the results we report models in which working memory capacity is defined using the two complex span measures of working memory capacity, and report the results of similar models using all three working memory capacity measures in the Appendix.

Table 9. *Exploratory factor analysis for Study 2.*

Measure	Factor 1	Factor 2	Factor 3	Factor 4
Stroop Squared	.17	.25	.24	.03
Flanker Squared	.32	.03	.55	-.11
Simon Squared	-.05	.55	.33	-.03
Antisaccade	.01	-.01	.64	-.05
FlankerDL	-.07	.24	-.24	.08
SACT	-.29	.04	.64	-.03
Selective Visual Arrays	.16	-.08	.53	.17
Raven's Matrices	.56	-.18	.14	.13
Letter Sets	.83	.20	-.26	-.08
Number Series	.86	.03	-.14	.00
Symmetry Span	.04	.04	-.14	.81
Rotation Span	-.06	.02	.02	.67
Mental Counters	.36	-.05	.35	.14
Digit Comparison	.08	.74	.03	-.06
Letter Comparison	.06	.74	-.16	.03
Pattern Comparison	-.01	.45	.15	.19
Eigenvalues	5.88	1.38	1.16	1.08

Note. Principal axis factoring with promax (oblique) rotation. Boldface indicates the strongest loading for each measure as well as any substantial cross-loadings.

Table 10. *Correlations between factors from exploratory factor analysis for Study 2.*

Factor	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1	---			
Factor 2	.61	---		
Factor 3	.71	.58	---	
Factor 4	.61	.44	.65	---

Confirmatory Factor Analyses

Next, we conducted a series of confirmatory factor analyses. We tested a model in which all the attention control measures were specified to load on a common factor. The model is depicted in Figure 17 ($\chi^2(9) = 34.65, p < .001$; CFI = .930, TLI = .884, RMSEA = .105 90% CI [.069, .143], SRMR = .055). The three Squared tests had the highest loadings, ranging from .62

to .77. The other attention control measures had slightly lower loadings, on average: antisaccade (.58), SACT (.36), selective visual arrays (.61).

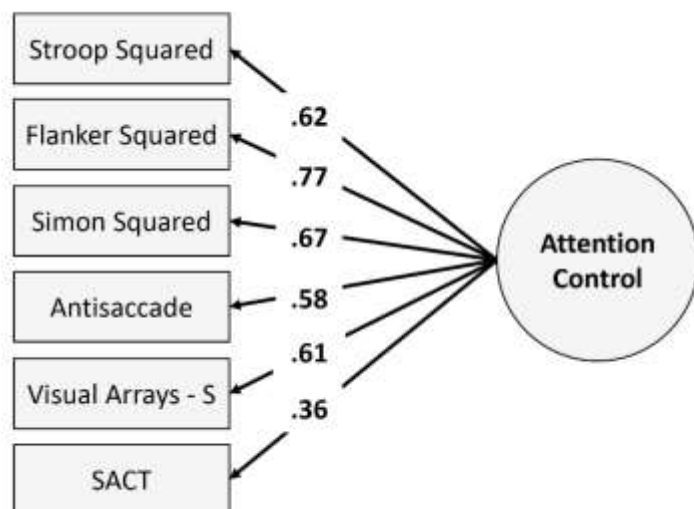


Figure 17. Latent variable model with all attention control measures loading on a common factor (Study 2).

Next, we specified a model in which the three Squared tests loaded on one factor and the remaining three attention control tests loaded on another factor. We allowed the two factors to correlate to determine how much variance they shared. The model is depicted in Figure 18 ($\chi^2(8) = 22.37, p = .004$; CFI = .961, TLI = .927, RMSEA = .083, 90% CI [.043, .125], SRMR = .042). The factor loadings for the three Squared tests ranged from .64 to .77, whereas the loadings for the antisaccade, SACT, and visual arrays tasks ranged from .43 to .68. The two factors were highly correlated, $r = .81, p < .001$, indicating that they shared 66% of their variance. Note the striking similarity to the online study results (i.e., the two factors correlated $r = .80, p < .001$). This provides further evidence for the construct validity of the three Squared tasks as measures of attention control. That said, setting the correlation between the latent factors equal to 1

resulted in significantly worse model fit, $\Delta\chi(1) = 12.28$, $p < .001$, indicating that a significant proportion of variance was unshared across the two sets of attention control tests.

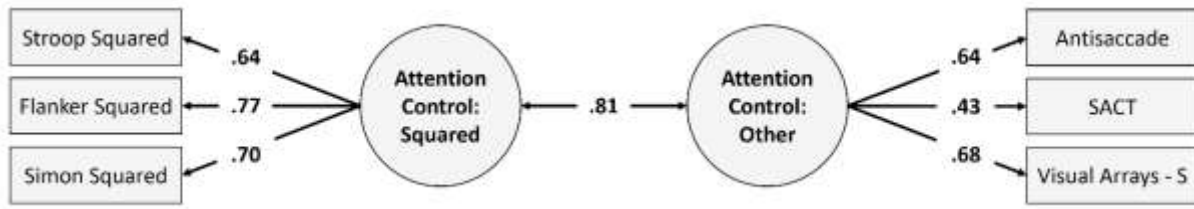


Figure 18. Latent variable model with the three Squared tests loading on one factor and the other attention control tests loading on another factor (Study 2).

In our next analyses we examined correlations between attention control, working memory capacity, fluid intelligence, and processing speed at the latent level. We created two attention control latent factors—one for the three Squared tests and another for the other tests of attention control—and correlated each with the other cognitive ability factors. The purpose of this model was to examine how correlations between attention control and the other latent cognitive ability factors differed depending on how attention control was measured. The fit of the model was good ($\chi^2(67) = 137.54$, $p < .001$; CFI = .934, TLI = .910, RMSEA = .067, 90% CI [.051, .083], SRMR = .054).

As shown in Table 11, the Squared attention control factor correlated $r = .77$ with the other attention control factor—note that in prior analyses the correlation was $r = .81$ (see Figure 18), but in this analysis the effective sample differs due to the inclusion of additional measures. The Squared attention control factor correlated $r = .71$ with fluid intelligence, whereas the other attention control factor correlated $r = .61$ with fluid intelligence. The difference was not statistically significant; we tested this by constraining the correlation between the other attention control factor and fluid intelligence to the same constant (i.e., “x”) as the correlation between the Squared attention control factor and fluid intelligence. Imposing this constraint did not

significantly worsen model fit, $\Delta\chi^2(1) = 1.89, p = .17$. The Squared attention control factor correlated $r = .52$ with the complex span working memory capacity factor, whereas the other attention control factor correlated $r = .61$ with working memory capacity. Once again, this difference was not statistically significant, $\Delta\chi^2(1) = 1.36, p = .24$. Finally, the Squared attention control factor correlated $r = .76$ with processing speed, whereas the other attention control factor correlated $r = .60$. The difference in these correlations was statistically significant, $\Delta\chi^2(1) = 4.62, p = .03$. Thus, the Squared attention control latent factor had a significantly stronger relationship with processing speed than did the other attention control latent factor, but the differences in correlations with fluid intelligence and working memory capacity were not statistically significant.

Table 11. *Latent variable correlations between the Squared attention control factor, the other attention control factor, fluid intelligence, working memory capacity, and processing speed.*

Factor	1.	2.	3.	4.	5.
1. Squared Attention Control	---				
2. Other Attention Control	.76	---			
3. Fluid Intelligence	.73	.60	---		
4. Working Memory Capacity	.68	.88	.78	---	
5. Processing Speed	.77	.59	.68	.60	---

Accounting for the Positive Manifold

In our next analyses, we investigated whether attention control could account for the substantial positive correlations observed among the cognitive ability factors. We tested two models, one in which we defined attention control using the three Squared tasks and another in which we used the other attention control measures. In both models, attention control was specified as a predictor of fluid intelligence, complex span working memory capacity, and processing speed. The residual variance in fluid intelligence, working memory capacity, and processing speed—that is, the variance that remained unaccounted for by attention control—was

allowed to correlate. The purpose of these analyses was to determine the extent to which partialing out variance in attention control reduced the latent correlation between cognitive ability factors. If the residual correlations are reduced to a considerable degree, or to non-significance, this would provide evidence that attention control captures domain-general variance that is shared by a number of different cognitive constructs.

The model using the three Squared tasks is depicted in Figure 19 ($\chi^2(38) = 103.67, p < .001$; CFI = .931, TLI = .900, RMSEA = .083, 90% CI [.064, .102], SRMR = .057). Squared attention control explained significant variance in each of the cognitive ability factors, with a standardized path of $\beta = .71$ ($p < .001$) to fluid intelligence, $\beta = .53$ ($p < .001$) to working memory capacity, and $\beta = .75$ ($p < .001$) to processing speed. The residual correlations between the cognitive ability factors were numerically lower after accounting for attention control. Residual fluid intelligence correlated $r = .28, p = .008$ with residual working memory capacity (reduced significantly from $r = .78$, see Table 11; $\Delta\chi^2(1) = 30.58, p < .001$); residual fluid intelligence correlated $r = .29, p = .011$ with residual processing speed (reduced significantly from $r = .68$, see Table 11; $\Delta\chi^2(1) = 15.62, p < .001$); residual working memory capacity correlated non-significantly ($r = .13, p = .260$) with residual processing speed (reduced significantly from $r = .60$, see Table 11; $\Delta\chi^2(1) = 19.55, p < .001$). These results indicate that the Squared attention control factor partly explains the covariation between fluid intelligence and other cognitive abilities, and fully explains the covariation between working memory capacity and processing speed.

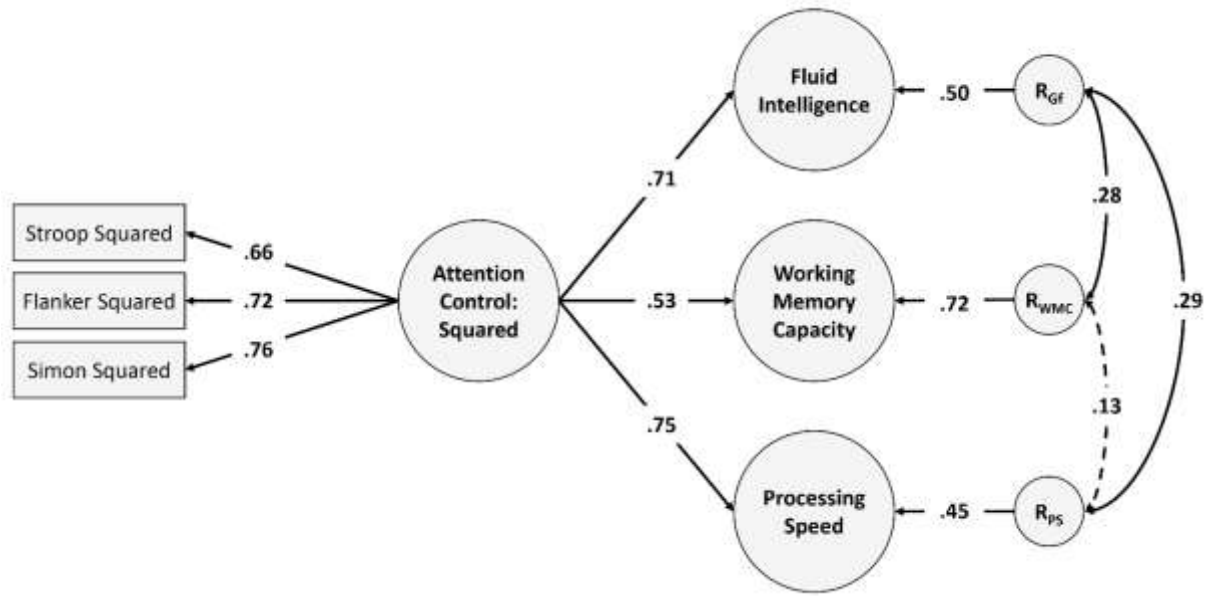


Figure 19. Structural equation model with a Squared attention control factor predicting fluid intelligence, complex span working memory capacity, and processing speed. The residual variance in each cognitive ability construct represents the variance in each construct after accounting for attention control. Indicators for fluid intelligence, working memory capacity, and processing speed are not depicted for visual clarity (Study 2).

The model using the three non-Squared tasks is depicted in Figure 20 ($\chi^2(38) = 82.01, p < .001$; CFI = .943, TLI = .917, RMSEA = .068, 90% CI [.047, .088], SRMR = .054). The non-Squared attention control factor explained significant variance in each of the cognitive ability factors, with a standardized path of $\beta = .62$ ($p < .001$) to fluid intelligence, $\beta = .61$ ($p = .003$) to working memory capacity, and $\beta = .61$ ($p < .001$) to processing speed. Residual fluid intelligence correlated $r = .26, p = .023$ with residual working memory capacity (reduced significantly from $r = .78$, see Table 11; $\Delta\chi^2(1) = 31.08, p < .001$); residual fluid intelligence correlated $r = .47, p < .001$ with residual processing speed (reduced significantly from $r = .68$, see Table 11; $\Delta\chi^2(1) = 7.05, p = .008$); residual working memory capacity correlated non-significantly ($r = .16, p = .176$) with residual processing speed (reduced significantly from $r = .60$, see Table 11; $\Delta\chi^2(1) =$

19.39, $p < .001$). In other words, the non-Squared attention control factor accounted for a significant portion of the positive manifold.

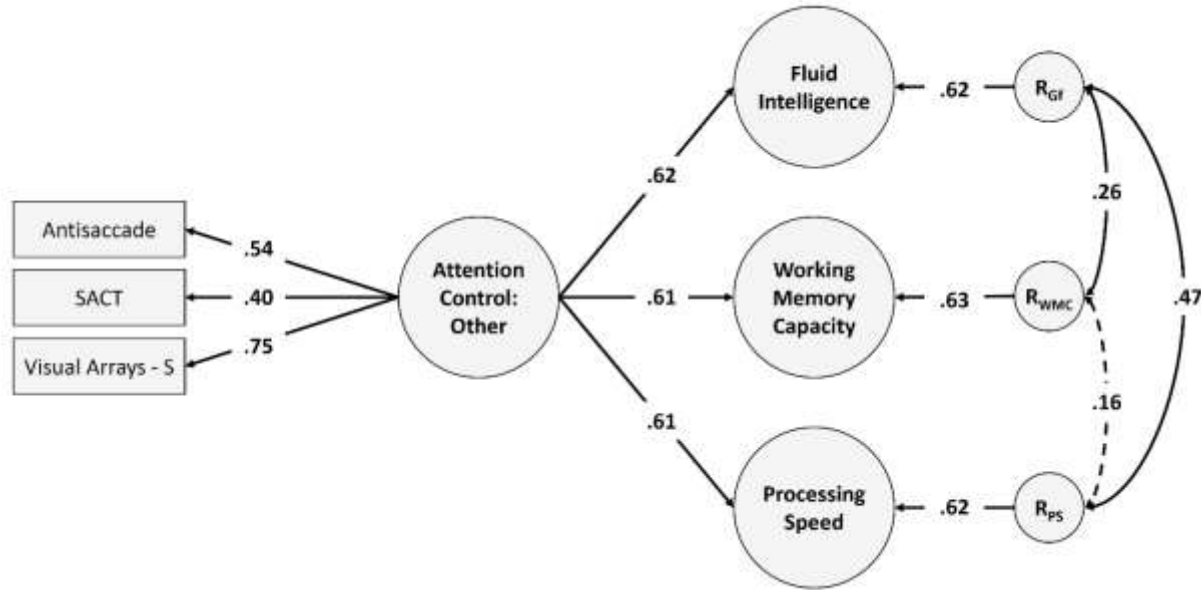


Figure 20. Structural equation model with a non-Squared attention control factor predicting fluid intelligence, working memory capacity, and processing speed. The residual variance in each cognitive ability construct represents the variance in each construct after accounting for attention control. Indicators for fluid intelligence, working memory capacity, and processing speed are not depicted for visual clarity.

Finally, we investigated whether processing speed could account for the positive correlations observed among the cognitive ability factors. Processing speed was specified as a predictor of attention control, fluid intelligence, and complex span working memory capacity. Attention control was defined using all six indicators. The residuals of the cognitive ability factors were allowed to correlate. The model is depicted in Figure 21 ($\chi^2(71) = 158.03, p < .001$; CFI = .918, TLI = .895, RMSEA = .072, 90% CI [.057, .087], SRMR = .058). Processing speed explained significant variance in each of the cognitive ability factors, with a standardized path of $\beta = .75$ ($p < .001$) to attention control, $\beta = .68$ ($p < .001$) to fluid intelligence, and $\beta = .45$ ($p < .001$) to working memory capacity. Residual attention control correlated $r = .46, p < .001$ with

residual fluid intelligence; residual attention control correlated $r = .40, p < .001$ with residual working memory capacity; residual fluid intelligence correlated $r = .34, p < .001$ with residual working memory capacity. Thus, processing speed accounted for a portion of the positive manifold but did not reduce any of the residual correlations between cognitive ability factors to non-significance.

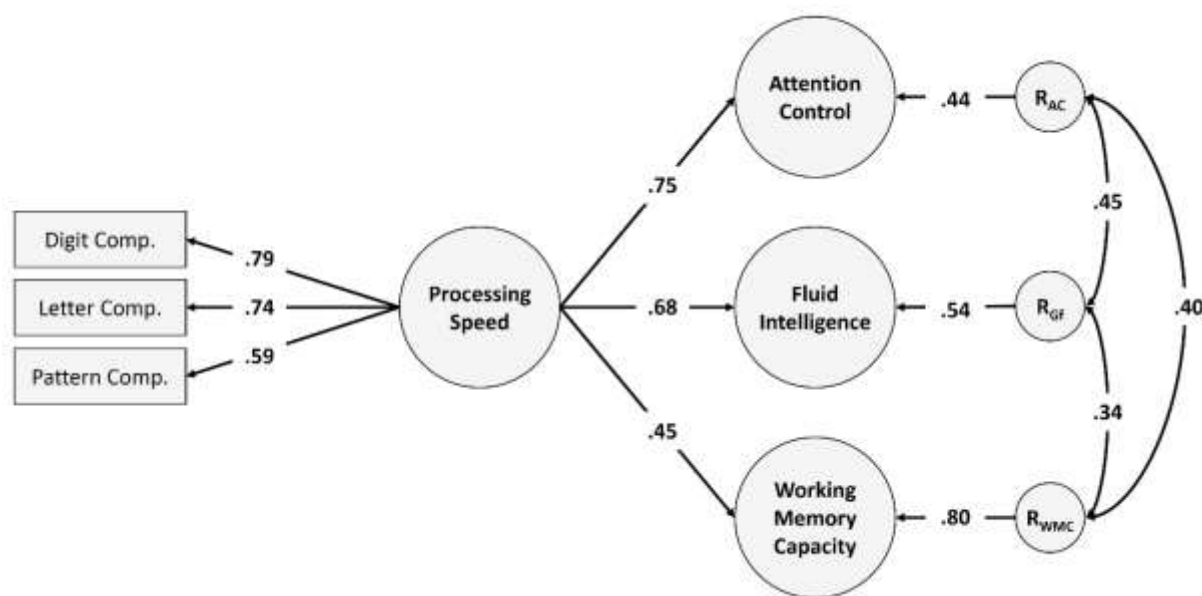


Figure 21. Structural equation model with a processing speed factor predicting attention control, fluid intelligence, and working memory capacity. The residual variance in each cognitive ability construct represents the variance in each construct after accounting for processing speed. Indicators for attention control, fluid intelligence, working memory capacity are not depicted for visual clarity.

Predicting Multitasking Ability

In this final section of analyses, we examine the relative contributions of different cognitive ability factors to multitasking ability. Our multitasking factor included four observed measures from three paradigms (SynWin, the Foster Multitask, and Control Tower: Primary and Distractor scores). These multitasking paradigms challenge participants to manage multiple

information processing demands simultaneously (or concurrently), including elements of visual and auditory processing, arithmetic, memory, symbol substitution, telling time, and problem solving. Thus, the multitasking factor extracted from these measures likely captures many different aspects of complex cognition, and in this case, serves as a proxy for real-world work performance. First, we tested whether a latent factor comprising just the Squared tests of attention control could explain variance in multitasking ability, and then repeated the analysis using the non-Squared tests of attention control as a point of comparison.

The Squared attention control factor had a standardized path of $\beta = .87$ to multitasking ability, indicating that it accounted for 75.6% of the variance in multitasking ability. The model is depicted in Figure 22 ($\chi^2 (13) = 27.95, p = .009$; CFI = .975, TLI = .959, RMSEA = .067, 90% CI [.032, .102], SRMR = .037).

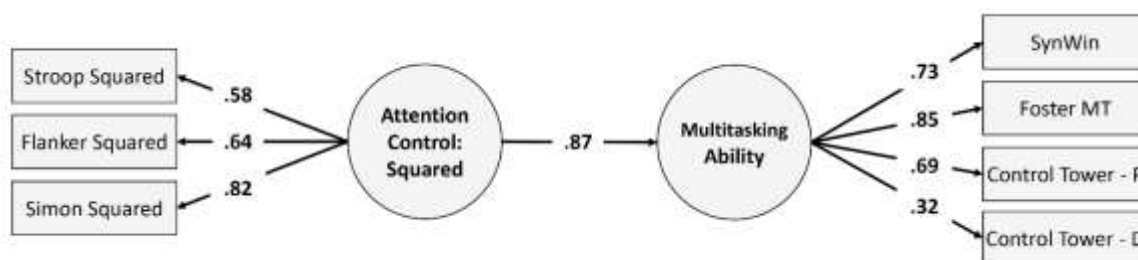


Figure 22. Structural equation model with a Squared attention control factor predicting multitasking ability.

For comparison, an attention control factor based on the non-Squared tests had a standardized path of $\beta = .75$ to multitasking ability, indicating that it accounted for 55.8% of the variance in multitasking ability. This model is depicted in Figure 23 ($\chi^2 (13) = 17.32, p = .185$; CFI = .990, TLI = .983, RMSEA = .036, 90% CI [.000, .076], SRMR = .037).

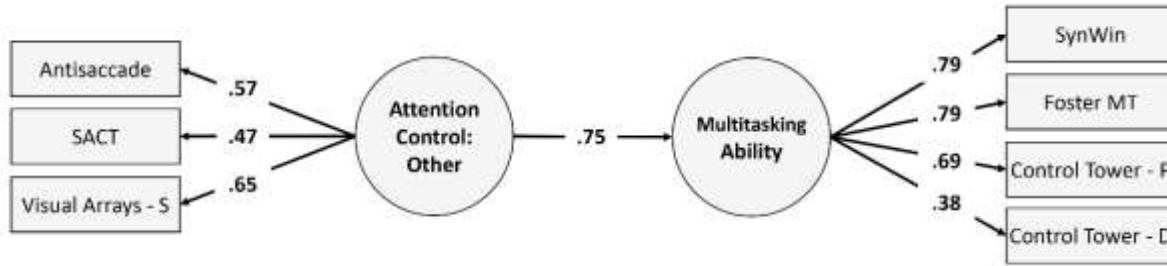


Figure 23. Structural equation model with a non-Squared attention control factor predicting multitasking ability.

Next, we tested a model in which both attention control factors were allowed to correlate and specified as predictors of multitasking ability. This analysis allows us to determine the relative contribution of each latent attention control factor while accounting for their covariation. The model is depicted in Figure 24 ($\chi^2(32) = 64.41, p < .001$; CFI = .950, TLI = .930, RMSEA = .069, 90% CI [.047, .092], SRMR = .049). The predictive path from Squared attention control to multitasking ability was substantial ($\beta = .69$), whereas the path from the other attention control factor to multitasking ability was smaller ($\beta = .23$). Setting the predictive paths equal to the same constant (i.e., “x”) significantly worsened model fit, $\Delta\chi^2(1) = 6.95, p = .008$, indicating that the Squared attention control factor and the other attention control factor differentially predicted multitasking performance. Although the two attention control factors share most of their reliable variance ($r = .75$), the results of these analyses indicate that the Squared attention control factor predicts multitasking above and beyond the other attention control factor. Combined, the two attention control factors accounted for 76.8% of the variance in multitasking ability.

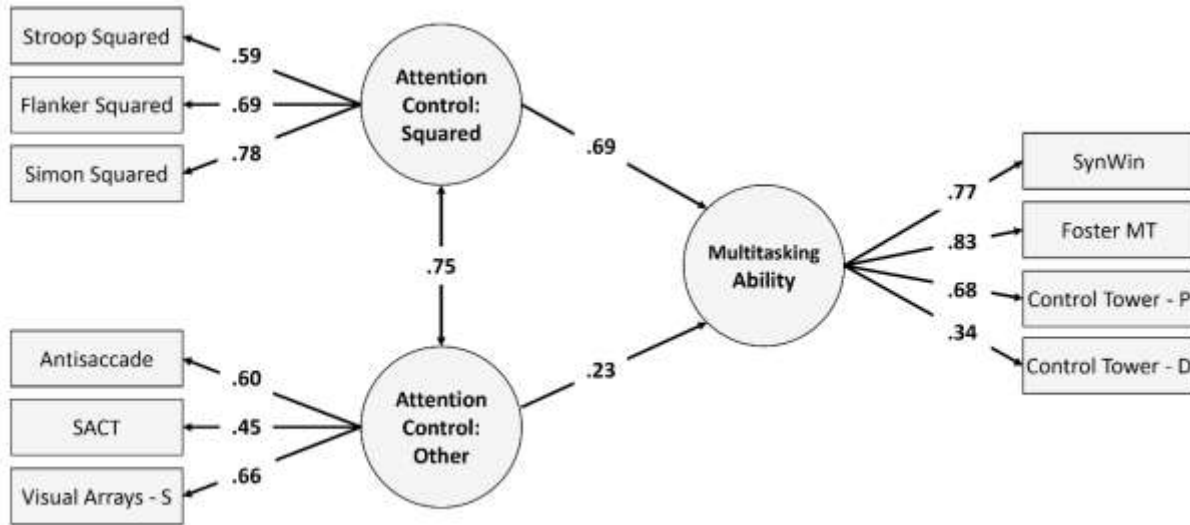


Figure 24. Structural equation model with the Squared attention control factor and the other attention control factor predicting multitasking ability.

Last, we tested a model in which latent factors representing attention control, fluid intelligence, working memory capacity, and processing speed were all specified as predictors of multitasking ability. The predictor factors were allowed to correlate, allowing us to determine the relative contribution of each cognitive ability factor above and beyond the other factors. We tested two versions of this model, one with the Squared tasks and one with the other attention control tasks.

The model using the Squared tasks is shown in Figure 25 ($\chi^2(80) = 163.43, p < .001$; CFI = .938, TLI = .918, RMSEA = .068, 90% CI [.053, .082], SRMR = .057). Fluid intelligence had the largest standardized path to multitasking ($\beta = .55$), followed by attention control ($\beta = .32$), processing speed ($\beta = .21$), and working memory capacity ($\beta = .07$). Combined, the predictors accounted for 100% of the variance in multitasking. In other words, individual differences in the ability to multitask were fully explained by a combination of fluid intelligence, attention control, processing speed, and, to a lesser extent, working memory capacity.

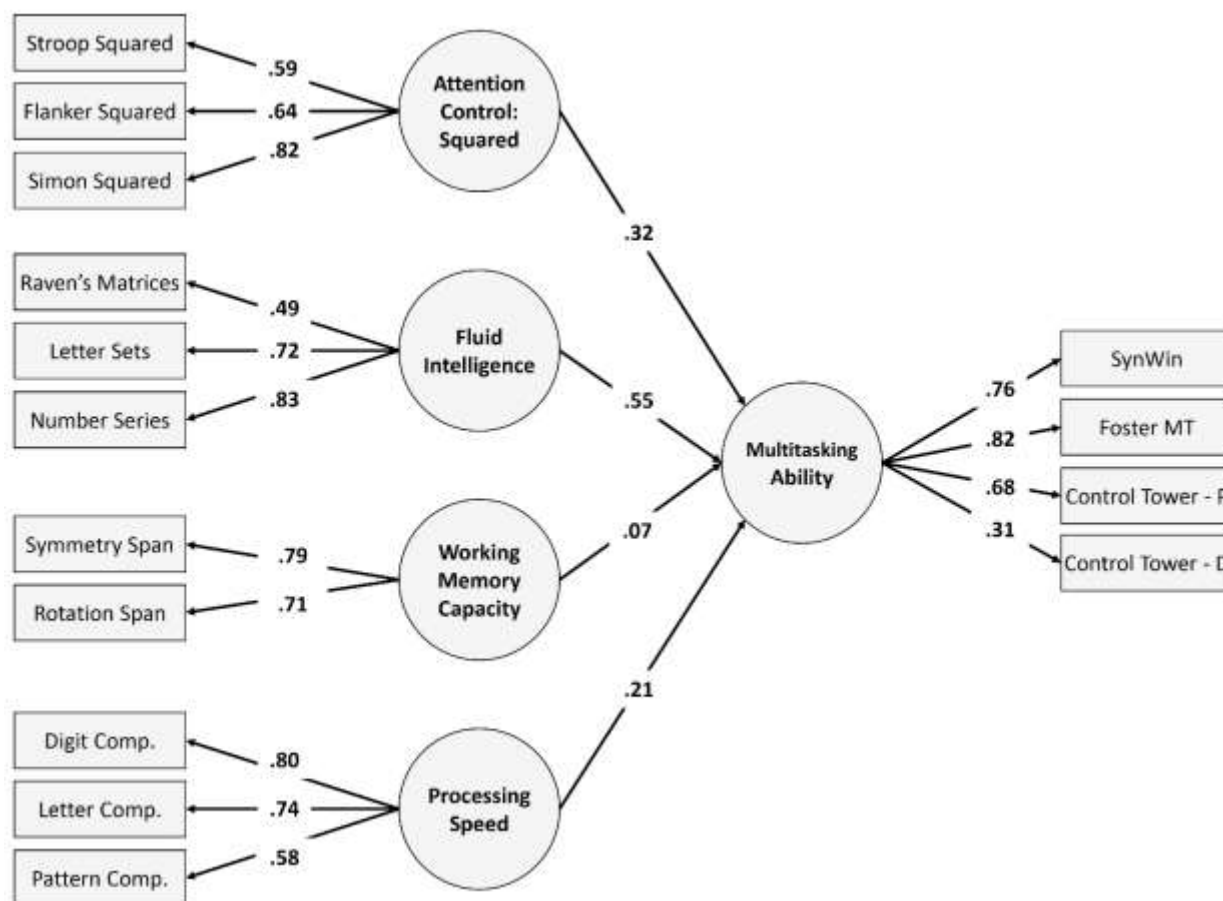


Figure 25. Structural equation model with Squared attention control, fluid intelligence, working memory capacity, and processing speed predicting multitasking ability. The cognitive ability factors were allowed to correlate, but the correlations are not shown here for visual clarity (AC with Gf, $r = .63$; AC with WMC, $r = .50$; AC with PS, $r = .75$; Gf with WMC, $r = .50$, Gf with PS, $r = .64$; WMC with PS, $r = .44$)

Finally, we tested an identical model to the previous one except we used the three non-Squared tasks as indicators of attention control. The model is shown in Figure 26 ($\chi^2(80) = 144.95, p < .001$; CFI = .944, TLI = .927, RMSEA = .060, 90% CI [.044, .075], SRMR = .055). Fluid intelligence had the largest standardized path to multitasking ($\beta = .56$), followed by processing speed ($\beta = .36$), non-squared attention control ($\beta = .23$), and working memory

capacity ($\beta = -.01$). Combined, the predictors accounted for 100% of the variance in multitasking.

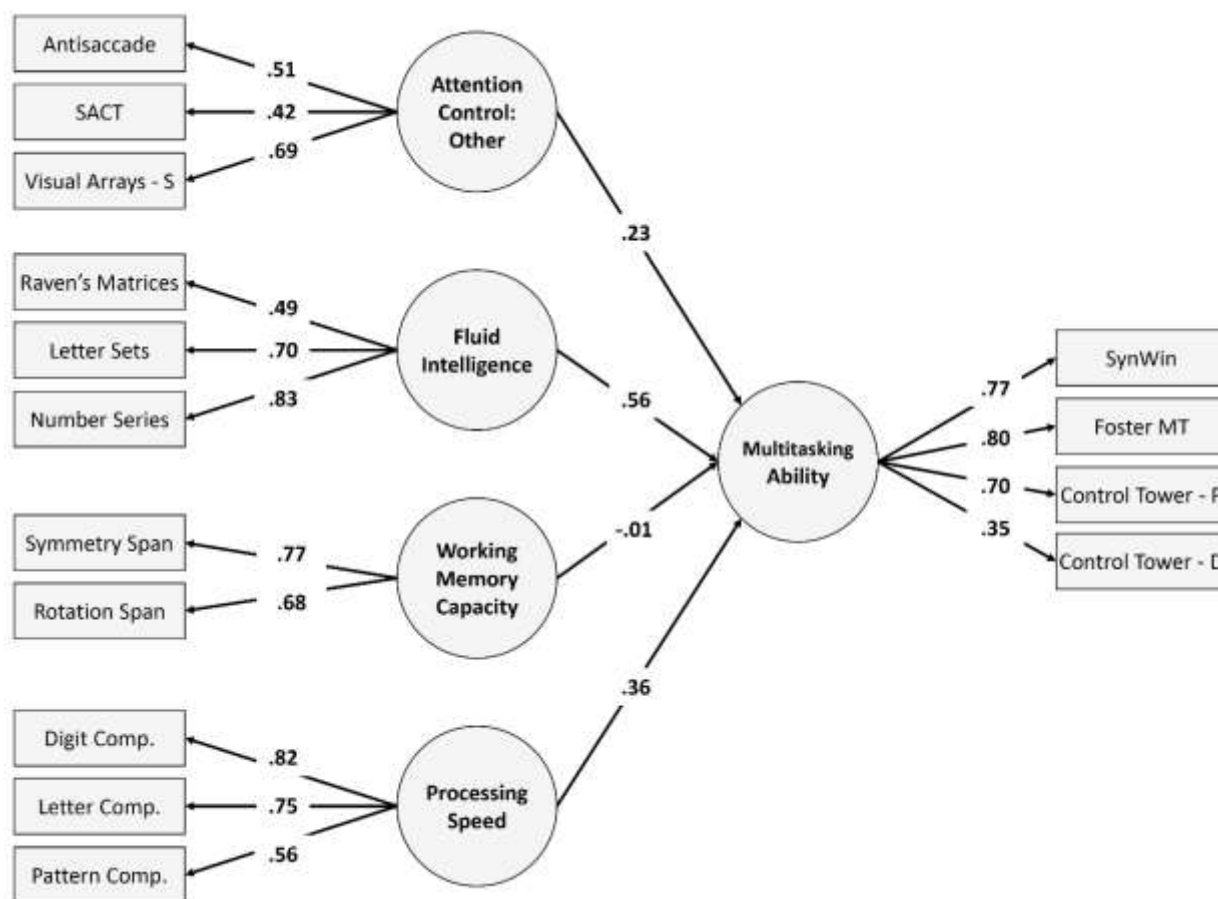


Figure 26. Structural equation model with non-Squared attention control, fluid intelligence, working memory capacity, and processing speed predicting multitasking ability. The cognitive ability factors were allowed to correlate, but the correlations are not shown here for visual clarity (AC with Gf, $r = .60$; AC with WMC, $r = .66$; AC with PS, $r = .56$; Gf with WMC, $r = .49$, Gf with PS, $r = .62$; WMC with PS, $r = .44$)

Analysis of Trial Types

As in Study 1, we found that participants earned the most points on Fully Congruent trials, and repeated measures ANOVAs revealed significant main effects of stimulus congruency

and response congruency and their interaction for everything except Simon Squared. For Simon Squared, we found non-significant main effects of stimulus congruency and response option congruency, but a significant interaction term. There was no other consistent pattern of differences in performance on the trial types of the three Squared tasks, and no consistent pattern of differences in correlations between performance on the trial types and other cognitive abilities. See Supplemental Materials for analyses.

General Discussion

To understand the nature of attention control as a cognitive construct, we need tasks with strong psychometric properties that produce systematic differences in performance across individuals. Measurement and theory are entwined; without adequate measurement, theoretical conclusions rest on tenuous ground. The purpose of this paper was to shed light on individual differences in attention control at the latent level by developing three new tests of attention control: Stroop Squared, Flanker Squared, and Simon Squared. We compared the psychometric properties and theoretical implications resulting from the use of these tasks with the best tasks to emerge from our lab's recent "toolbox approach" to improving the measurement of attention control (Draheim et al., 2021).

Internal Consistency and Test-Retest Reliability

The three Squared tasks had very high internal consistency estimates. In Study 1, split-half reliability estimates ranged from .93 to .97, and in Study 2, they ranged from .94 to .97. By comparison, the best tasks to emerge from our lab's "toolbox" paper had internal consistency estimates ranging from .58 to .95 in Study 1 and from .87 to .91 in Study 2. In Study 2, we administered the Squared tasks three times, twice in the lab and once as a follow-up test which was completed on participants' computers outside the lab. We found test-retest reliabilities

ranging from $r = .53$ to $r = .75$ for the first and second test administrations of the Squared tasks, and from $r = .46$ to $r = .55$ for the second and third administrations. Correspondingly, we found very small practice effects on the Squared tasks; if anything, participants performed slightly worse on subsequent attempts (Figure 16), but changes in performance were generally not statistically significant.

Convergent Validity and Construct Validity

The Squared tasks demonstrated convergent validity and appear to reflect individual differences in attention control. At the observed level, the Squared tasks had strong intercorrelations, with an average of $r = .51$ for Study 1 and $r = .50$ for Study 2. At the latent level, the Squared tasks had the highest loadings on a common attention control factor that included other attention control measures and demonstrated good model fit. When we specified two attention control factors, one for the Squared tasks and one for the other attention control tasks, we found that the two factors correlated $r = .80$ in Study 1 and $r = .81$ in Study 2. This indicates that the Squared tasks share a majority of their reliable variance with the other measures of attention control used in these studies. Nevertheless, the Squared tasks did capture some unique variance that set them apart, precluding a perfect correlation between the latent factors without significantly compromising model fit.

Accounting for the Positive Manifold

In both studies, we found that the Squared attention control tasks accounted for a significant proportion of the covariation between fluid intelligence and working memory, but did not reduce the residual correlation between these constructs to zero. We found a similar pattern of results when using the other attention control tasks. This provides further evidence for the executive attention view, which argues that the primary “active ingredient” tapped by working

memory capacity measures that explains the correlation between working memory capacity and fluid intelligence is attention control. Nevertheless, the statistically significant residual correlation points to other factors beyond the ability to control attention that may contribute to this relationship. For example, retrieval from secondary memory may also play a role (Unsworth et al., 2014).

In Study 2, we found that attention control fully explained the correlation between working memory capacity and processing speed, regardless of whether it was measured using the Squared tasks or the other attention control tests. We also found that attention control explained most of the covariance between fluid intelligence and processing speed, but did not eliminate it. Comparing the Squared tasks to the other attention control tasks, we found that the Squared tasks had a significantly stronger relationship with processing speed. This could be due to the speeded component of the Squared tasks, which is shared with processing speed tests: participants earn points by correctly responding to as many trials as they can within a fixed time limit. It is possible that the speed component tapped by the Squared tasks is the reason why, at the latent level, the Squared tasks and the other attention control tasks did not correlate perfectly.

The broader purpose of these analyses was to determine the extent to which attention control explains the *positive manifold*—the positive correlations observed among broad cognitive abilities. We have argued that attention control is a domain-general ability that is required by a wide range of cognitive tasks, helping to explain why individuals who perform below average on one cognitive test tend to perform below average on other cognitive tests, too (Burgoyne et al., 2022). In support of this view, attention control accounted for a significant portion of the covariation between all of the broad cognitive abilities we measured (i.e., fluid intelligence, working memory capacity, and processing speed). For comparison, processing speed accounted

for a small portion of the covariation between the other cognitive ability constructs, and did not fully account for any of them (Figure 21). This suggests that attention control may be more fundamental to explaining the positive manifold than processing speed, although we note that this is a contentious issue that will require multiple convergent methods to substantiate.

Predicting Multitasking Ability

Again, we found that multitasking, reflected by the tasks used here, constitutes a coherent latent construct. Which abilities are important to explaining individual differences in multitasking? On their own, the three Squared tasks explained 75% of the variance in multitasking ability at the latent level, whereas the other attention control tasks explained around 55% of the variance. When the two attention control factors were modeled as joint predictors of multitasking ability, the predictive path from the Squared attention control factor was significantly larger than the path from the other attention control factor (β s of .69 vs. .23). Despite these differences in predictive validity, in general, attention control appears to play a critical role in the ability to effectively manage multiple task demands simultaneously (or concurrently). When we included other cognitive ability predictors in the model, we found that 100% of the variance in multitasking ability could be explained by a combination of fluid intelligence, attention control, processing speed, and to a lesser extent, working memory capacity. Multitasking is a complex cognitive ability that captures a range of information processing demands. It seems fitting, then, that a combination of factors, including not only the ability to control attention but also the ability to solve novel problems and process information quickly, contribute to individual differences in performance.

Administration Time

The average administration time for each of the three Squared tasks was 2 minutes, amounting to 6 minutes of total testing time for the average participant. For comparison, the best three tasks from our lab's "toolbox" paper each required 12.5 minutes, on average, amounting to 37.5 minutes of total testing time for the average participant. Considering that the Squared tasks accounted for 20% more variance in multitasking ability, 24% more variance in processing speed, and accounted for as much of the covariation between cognitive abilities as the other attention control tests did, the potential savings in time costs associated with the Squared tasks is substantial. In less than ten minutes, researchers can obtain three reliable and valid measures of attention control with strong loadings on a common factor, permitting analyses and conclusions at the level of latent cognitive constructs instead of at the level of observed measures.

Furthermore, the tests can easily be administered on participants' own computers or online. From a practical perspective, the three Squared tests of attention control will allow researchers to conduct more extensive studies of individual differences in cognitive abilities by sparing time for the measurement of other constructs.

The Nature and Measurement of Attention Control

There is perhaps no field in psychology where advances in theory, quantitative methods, and measurement are so intimately interweaved as they are in differential psychology. This is as true today (Burgoyne et al., 2022; Draheim et al., 2021) as it was in the early days of intelligence research. For instance, the invention of quantitative methods, such as the correlation statistic, was driven by the need to quantify the relation between various tests of mental ability (Galton, 1889; Spearman, 1904). The success of the correlation statistic led to the creation of a more

diverse set of mental ability tests, the development of factor analytic methods, and standardized testing—all of which were both motivated by and informed advances in theories of intelligence.

This interweaving of theory, quantitative methods, and measurement continued throughout the 20th and 21st centuries in many domains of differential psychology. In cognitive psychology, we witnessed the blending of new concepts such as working memory and executive attention in the experimental tradition and the development of novel measures of simple and complex memory span in the differential tradition. This blending led to the concept of individual differences in working memory capacity and the role of executive attention in memory and other cognitive abilities. Although research on individual differences in working memory capacity became highly influential in how we think about cognitive abilities (Burgoyne et al., 2022; Burgoyne & Engle, 2020; Engle, 2018; Engle, 2002), it has not been without controversy.

The controversy we find ourselves in today concerns the nature and measurement of attention control. We have argued that the core ingredient in measures of working memory capacity is the domain-general control of attention — “the *capacity for controlled, sustained attention in the face of interference or distraction*... working memory capacity reflects the ability to apply activation to memory representations, to either bring them into focus or maintain them in focus, particularly in the face of interference or distraction” (Engle et al. 1999, italics added). As such, measures of working memory capacity have long been used as a proxy measure for this domain-general ability to control attention.

As our theories about the nature of attention control developed, however, there was a need to measure attention control directly with tasks that did not emphasize short-term memory demands. A natural place to look for such tasks was the experimental tradition, as there was already a large body of research on attention, distractor interference, and conflict resolution. In

some ways, borrowing tasks from the experimental tradition was largely successful (Miyake & Friedman, 2012; Redick et al., 2016). In less obvious ways, there was a measurement problem that has now led researchers to question whether we should even think of attention control as an individual differences construct (Rey-Mermet et al., 2018). Therefore, differential psychology finds itself once again at a pivotal moment where theory, quantitative methods, and measurement are entwined and will likely determine the future of research on individual differences in cognitive ability.

Conclusion

Our position is that individual differences in the ability to control attention can be reliably measured and they underpin a wide range of cognitive functions, from problem solving and storing information in working memory to processing information rapidly and multitasking. That said, it is critical that we continue to refine our tools, including not only our tasks but also our experimental and statistical approaches. In this paper, we demonstrated that three “Squared” tests of attention control can provide an efficient, reliable, and valid estimate of individual differences in the ability to control attention. We hope that these new tools will prove fruitful to researchers interested in advancing scientific understanding of attention control.

Context

Reliably measuring individual differences in the ability to control attention has posed a challenge for psychologists. The crux of the problem is that researchers have used experimental paradigms (e.g., Stroop) with poor psychometric properties when used for differential psychology, primarily because these tasks use response time difference scores. Unreliability attenuates correlations, which has led some researchers to accept the null hypothesis that attention control is not a coherent cognitive construct, and others to argue that it is unimportant

in explaining individual differences in real-world outcomes. Measurement and theory are entwined; for researchers to draw firm theoretical conclusions, they must have a solid methodological framework with reliable and valid measurement instruments for those arguments to rest on. To this end, we developed three efficient, reliable, and valid tests of attention control (Stroop Squared, Flanker Squared, and Simon Squared), and used them to examine how attention control relates to higher-order cognitive constructs as well as proxies for real-world performance. We found compelling evidence for a unitary attention control latent factor, which was highly correlated with fluid intelligence, working memory capacity, and processing speed, and helped explain their covariation. Furthermore, attention control explained a majority of the variance in multitasking ability.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131(1), 30-60.
- Ackerman, P. L., & Hambrick, D. Z. (2020). A primer on assessing intelligence in laboratory studies. *Intelligence*, 80, 101440.
- Ahmed, S. F., Tang, S., Waters, N. E., & Davis-Kean, P. (2019). Executive function and academic achievement: Longitudinal relations from early childhood to adolescence. *Journal of Educational Psychology*, 111(3), 446.
- Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT Battery, *Military Psychology*, 9, 5-37.
- Allan, J. L., McMinn, D., & Daly, M. (2016). A bidirectional relationship between executive function and health behavior: evidence, implications, and future directions. *Frontiers in neuroscience*, 10, 386.
- Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A*, 49, 5–28.
- Baumeister, R. F., Schmeichel, B. J., & Vohs, K. D. (2007). Self-regulation and the executive function: The self as controlling agent. *Social psychology: Handbook of basic principles*, 2, 516-539.
- Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and individual differences*, 21(4), 327-336.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–652.

- Burgoyne, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science*, 29(6), 624-630.
- Burgoyne, A. P., Hambrick, D. Z., & Altmann, E. M. (2021). Incremental validity of placekeeping as a predictor of multitasking. *Psychological Research*, 85(4), 1515-1528.
- Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., & Engle, R. W. (2022). Attention Control and Process Overlap Theory: Searching for Cognitive Processes Underpinning the Positive Manifold. *Intelligence*, 91, 101629.
- Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., Hambrick, D. Z., & Engle, R. W. (2021). Understanding the relationship between rationality and intelligence: A latent-variable approach. *Thinking and Reasoning*, 1-42.
- Burgoyne, A. P., Hambrick, D. Z., & Altmann, E. M. (2019). Is working memory capacity a causal factor in fluid intelligence? *Psychonomic Bulletin & Review*, 26(4), 1333-1339.
- Burgoyne, A. P., Tsukahara, J. S., Draheim, C., & Engle, R. W. (2020). Differential and experimental approaches to studying intelligence in humans and non-human animals. *Learning and Motivation*, 72, 101689.
- Chiou, J. S., & Spreng, R. A. (1996). The reliability of difference scores: A re-examination. *Journal of Consumer Satisfaction Dissatisfaction and Complaining Behavior*, 9, 158-167.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30(2), 163-184. [https://doi.org/10.1016/S0160-2896\(01\)00096-4](https://doi.org/10.1016/S0160-2896(01)00096-4)

- Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51, 42–100.
<http://dx.doi.org/10.1016/j.cogpsych.2004.12.001>
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin*, 74(1), 68–80. <https://doi.org/10.1037/h0029382>
- Diamond, A. (2013). Executive functions. *Annual review of psychology*, 64, 135-168.
- Donders, F. C. (1868). Over de snelheid van psychische processen. *Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool (1968–1869)*, 2, 92-120.
- Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task switching as a case example. *Perspectives on Psychological Science*, 11(1), 133-155.
- Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, 145(5), 508–535. <https://doi.org/10.1037/bul0000192>
- Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, 150(2), 242.
- Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review*.
<https://doi.org/10.3758/s13423-021-02052-2>

- Ekstrom R. B., French J. W., Harman H. H., & Dermen D. (1976). *Manual for kit of factor-referenced cognitive tests: 1976*. Educational Testing Service.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19–23.
- Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, 13(2), 190-193.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331. <https://doi.org/10.1037/0096-3445.128.3.309>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, 16(1), 143–149. <https://doi.org/10.3758/BF03203267>
- Friedman, N. P. (2019). Research on individual differences in executive functions. *Bilingualism, Executive Function, and Beyond: Questions and insights*, 57, 209.
- Friedman, N. P., & Miyake, A. (2004). The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology: General*, 133(1), 101–135. <https://doi.org/10.1037/0096-3445.133.1.101>
- Fukuda, K., Woodman, G. F., & Vogel, E. K. (2015). Individual differences in visual working memory capacity: Contributions of attentional control to storage. *Mechanisms of sensory working memory: Attention and performance XXV*, 105.
- Haaf, J. M., & Rouder, J. N. (2017). Developing constraint in bayesian mixed models. *Psychological Methods*, 22, 779-798.

- Hall, P. A., Fong, G. T., Epp, L. J., & Elias, L. J. (2008). Executive function moderates the intention-behavior link for physical activity and dietary behavior. *Psychology & Health, 23*(3), 309-326.
- Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18*, 1279–1296.
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166-1186.
- Hedge, C., Powell, G., Bompas, A., & Sumner, P. (2021). Strategy and processing speed eclipse individual differences in control ability in conflict tasks. *Journal of Experimental psychology. Learning, Memory, and Cognition*.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55.
- Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 645–662. <https://doi.org/10.1037/0278-7393.33.4.645>
- 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*(4), 851.
- Galton, F. (1889). I. Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London, 45*(273-279), 135-145.

- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49(3), 227-229.
- 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(4), 637-671.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of experimental psychology: General*, 132(1), 47.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131(1), 66–71. <https://doi.org/10.1037/0033-2909.131.1.66>
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, 133(2), 189–217. <https://doi.org/10.1037/0096-3445.133.2.189>
- Kofler, M. J., Soto, E. F., Fosco, W. D., Irwin, L. N., Wells, E. L., & Sarver, D. E. (2020). Working memory and information processing in ADHD: Evidence for directionality of effects. *Neuropsychology*, 34(2), 127.
- Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, 27(3), 151-177.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14(4), 389-433.

- Lerche, V., von Krause, M., Voss, A., Frischkorn, G. T., Schubert, A. L., & Hagemann, D. (2020). Diffusion modeling and intelligence: Drift rates show both domain-general and domain-specific relations with intelligence. *Journal of Experimental Psychology: General*, 149(12), 2207.
- Lezak, M. D. (1982). The problem of assessing executive functions. *International journal of Psychology*, 17(1-4), 281-297.
- Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance*, 5(2), 189–207. <https://doi.org/10.1037/0096-1523.5.2.189>
- Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657), 279-281.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. <https://doi.org/10.1037/0033-2909.109.2.163>
- Martin, J., Mashburn, C. A., & Engle, R. W. (2020a). Improving the Validity of the Armed Service Vocational Aptitude Battery with Measures of Attention Control. *Journal of Applied Research in Memory and Cognition*, 9(3), 323-335.
- Martin, J. D., Shipstead, Z., Harrison, T. L., Redick, T. S., Bunting, M., & Engle, R. W. (2020b). The role of maintenance and disengagement in predicting reading comprehension and vocabulary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(1), 140–154. <https://doi.org/10.1037/xlm0000705>

- Martin, J. D., Tsukahara, J. S., Draheim, C., Shipstead, Z., Mashburn, C. A., Vogel, E. K., & Engle, R. W. (2021). The visual arrays task: Visual storage capacity or attention control? *Journal of Experimental Psychology: General*. Advance online publication. <https://doi.org/10.1037/xge0001048>
- McCabe, D. P., Roediger III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: evidence for a common executive attention construct. *Neuropsychology*, 24(2), 222.
- 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(2), 302–320. <https://doi.org/10.1037/a0025250>
- 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(1), 8-14.
- 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(1), 49-100.
- Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working Memory and Intelligence--Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131(1), 61–65. <https://doi.org/10.1037/0033-2909.131.1.61>

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.

Psychology Software Tools, Inc. [E-Prime Go]. (2020). Retrieved from <https://support.pstnet.com/>.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human perception and performance*, 26(1), 127.

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales* (Vol. 759). Oxford: Oxford Psychologists Press.

Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., ... & 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(11), 1473.

Redick, T. S., Unsworth, N., Kelly, A. J., & Engle, R. W. (2012). Faster, smarter? Working memory capacity and perceptual speed in relation to fluid intelligence. *Journal of Cognitive Psychology*, 24(7), 844-854.

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(4), 501.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, 26(2), 452-467.

- Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental psychology*, 27(5), 763.
- Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin & Review*, 15(2), 364-371.
- Schmeichel, B. J., & Demaree, H. A. (2010). Working memory capacity and spontaneous emotion regulation: High capacity predicts self-enhancement in response to negative feedback. *Emotion*, 10(5), 739–744. <https://doi.org/10.1037/a0019355>
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609-612.
- Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, 11(6), 771-799.
- Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 51(3), 300–304. <https://doi.org/10.1037/h0020586>
- Spearman, C. (1904). “General Intelligence,” Objectively Determined and Measured. *The American Journal of Psychology*, 15(2), 201-292.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87, 245-251.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric monographs*.

- Tsukahara, J. S., Harrison, T. L., Draheim, C., Martin, J. D., & Engle, R. W. (2020). Attention control: The missing link between sensory discrimination and intelligence. *Attention, Perception, & Psychophysics*, 82(7), 3445-3478.
- 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.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498-505.
- Wiley, J., & Jarosz, A. F. (2012). *How working memory capacity affects problem solving*. In Psychology of learning and motivation (Vol. 56, pp. 185-227). Academic Press.
- Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2011). Contributions of modern measurement theory to measuring executive function in early childhood: An empirical demonstration. *Journal of experimental child psychology*, 108(3), 414-435.
- Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, 78(4), 16-33.
- Zelazo, P. D., & Cunningham, W. A. (2007). Executive Function: Mechanisms Underlying Emotion Regulation. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 135–158). The Guilford Press.

Supplemental Materials

Study 1: Analysis of Trial Types

In the following section, we conduct an exploratory analysis on the different trial types found in the Squared tasks. To reiterate, each Squared test of attention control has four kinds of trials which were sampled with equal probability. Our prediction was that “fully congruent” trials (i.e., the target stimulus and response options are congruent) would be the easiest for participants because they require the least amount of conflict resolution. By contrast, we predicted that “fully incongruent” trials (i.e., the target stimulus and response options are incongruent) would be the most difficult because they require the most amount of conflict resolution. We compared performance on the different trial types and correlated performance on each trial type with other measures of cognitive ability.

As shown in Figure S1, participants earned the most points and responded quickest on fully congruent trials in all three Squared tasks. Thus, fully congruent trials may be the easiest trial type, supporting our prediction. However, contrary to prediction, there was no pattern of performance differences across the other trial types that was consistent across the three Squared tasks. We ran a series of repeated measures ANOVAs to examine the effect of stimulus congruency, response option congruency, and their interaction on mean total points earned and mean response times (Table S1). We found significant main effects and interaction effects for everything except Simon Squared. For Simon Squared, we found non-significant main effects of stimulus congruency and response option congruency, but a significant interaction between stimulus congruency and response option congruency.

Table S1. Repeated-measures ANOVAs on the different trial types in the three Squared tasks.

	Main effect of Stimulus Congruency	Main effect of Response Congruency	Interaction Effect
Stroop Squared Points	$F(1, 310) = 74.47,$ $p < .001, \eta_p^2 = .194$	$F(1, 310) = 138.40,$ $p < .001, \eta_p^2 = .309$	$F(1, 310) = 55.26,$ $p < .001, \eta_p^2 = .151$

NATURE AND MEASUREMENT OF ATTENTION CONTROL

Flanker Squared Points	$F(1, 296) = 53.22,$ $p < .001, \eta_p^2 = .152$	$F(1, 296) = 4.74,$ $p = .03, \eta_p^2 = .016$	$F(1, 296) = 29.22,$ $p < .001, \eta_p^2 = .090$
Simon Squared Points	$F(1, 320) = 0.01,$ $p = .94, \eta_p^2 = .000$	$F(1, 320) = 0.69,$ $p = .41, \eta_p^2 = .002$	$F(1, 320) = 12.96,$ $p < .001, \eta_p^2 = .039$
Stroop Squared RT	$F(1, 310) = 17.43,$ $p < .001, \eta_p^2 = .053$	$F(1, 310) = 197.98,$ $p < .001, \eta_p^2 = .390$	$F(1, 310) = 14.39,$ $p < .001, \eta_p^2 = .044$
Flanker Squared RT	$F(1, 296) = 193.64,$ $p < .001, \eta_p^2 = .395$	$F(1, 296) = 234.44,$ $p < .001, \eta_p^2 = .442$	$F(1, 296) = 36.18,$ $p < .001, \eta_p^2 = .109$
Simon Squared RT	$F(1, 320) = 1.66,$ $p = .20, \eta_p^2 = .005$	$F(1, 320) = 0.02,$ $p = .89, \eta_p^2 = .000$	$F(1, 320) = 18.90,$ $p < .001, \eta_p^2 = .056$

Note: η_p^2 = partial eta squared, a measure of the effect size.

NATURE AND MEASUREMENT OF ATTENTION CONTROL

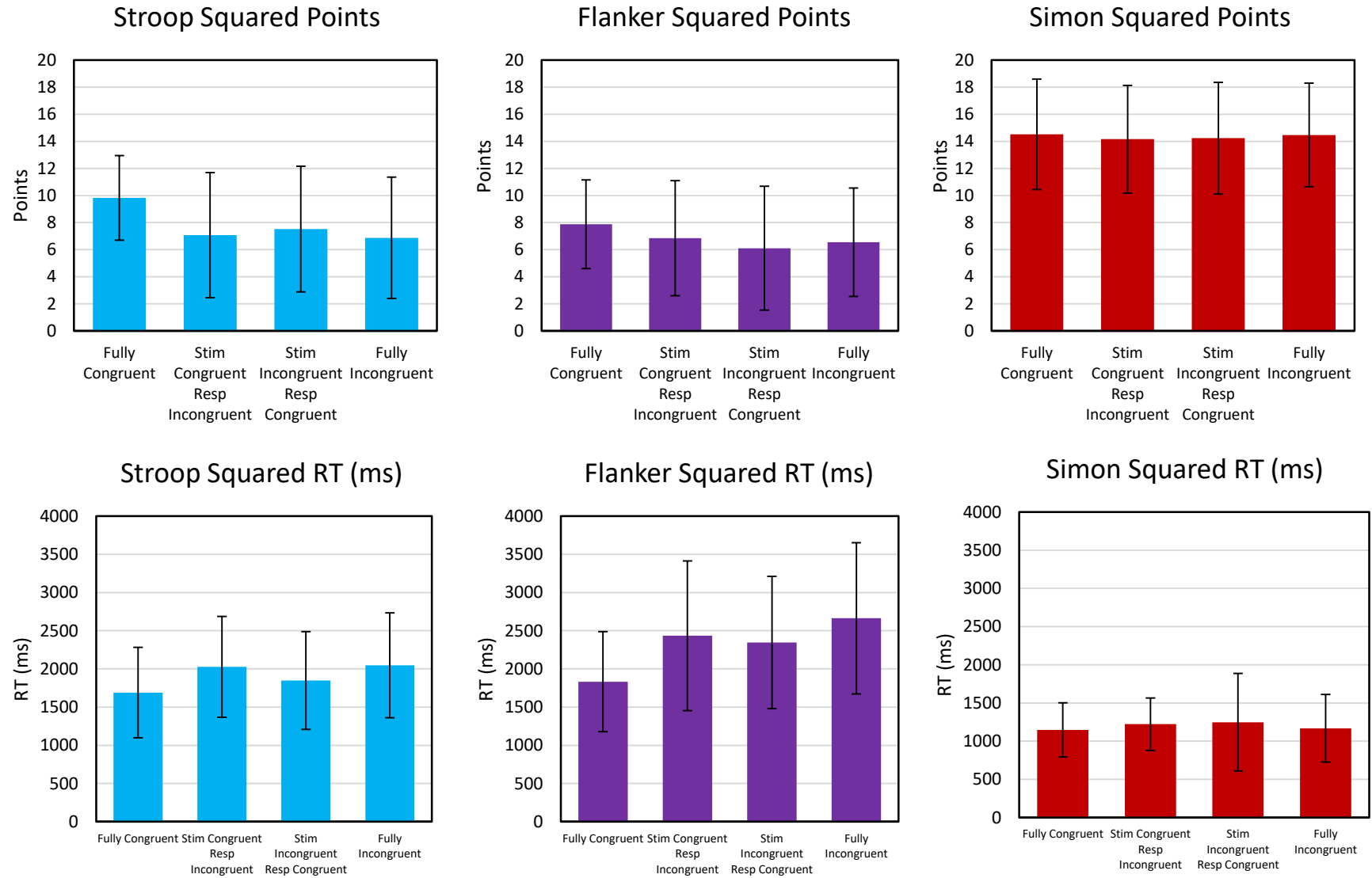


Figure S1. Performance on the four trial types in the Squared tasks. Error bars represent ± 1 SD (Study 1).

NATURE AND MEASUREMENT OF ATTENTION CONTROL

Next, we examined whether there were different patterns of correlations between performance on each of the trial types in the Squared tasks and the other cognitive ability factors (Table S2). We saved factor scores representing performance on the other attention control tests (i.e., antisaccade, SACT, and selective visual arrays), the fluid intelligence tests (i.e., Raven's matrices, letter sets, and number series), and the working memory capacity tests (i.e., symmetry span, rotation span, and mental counters) by performing three principal components analyses, one for each cognitive construct, and saving participants' scores on each factor. We used a similar approach for points earned and mean response times on each of the trial types in the Squared tasks. That is, we conducted eight principal components analyses, entering into each analysis the total points earned (or mean reaction time) for a given trial type for all three Squared tasks. For example, we entered total points earned on fully congruent trials in Stroop Squared, Flanker Squared, and Simon Squared into a principal components analysis and saved the resulting factor score (i.e., component score). We then examined correlations between each of these markers of performance on the different trial types and the cognitive ability component scores.

Table S2. *Correlations between performance on each trial type in the Squared tasks and cognitive ability component scores.*

Measure	Attention Control: Other	Fluid Intelligence	Working Memory Capacity
Fully Congruent Points	.49	.58	.31
Fully Incongruent Points	.50	.57	.40
Stimulus Congruent Response Incongruent Points	.51	.57	.39
Stimulus Incongruent Response Congruent Points	.48	.57	.31
Fully Congruent RT	-.45	-.54	-.30
Fully Incongruent RT	-.41	-.52	-.31
Stimulus Congruent Response Incongruent RT	-.48	-.51	-.29
Stimulus Incongruent Response Congruent RT	-.46	-.50	-.31

Note: *N* ranges from 223 to 262. Each variable in this analysis is a component score extracted from a principal components analysis; one for each trial type or cognitive ability measure.

For points earned, the correlations across the different Squared task trial types ranged from $r = .48$ to $r = .51$ with attention control, from $r = .57$ to $r = .58$ with fluid intelligence, and from $r = .31$ to $r = .40$ with working memory capacity. Thus, correlations with attention control and fluid intelligence were remarkably similar across trial types, but there was a slight difference in the magnitude of the correlations with working memory capacity. We performed Steiger's Z test (Steiger, 1980) to determine whether the difference between these dependent correlations was statistically significant. Indeed, fully incongruent points correlated significantly more strongly with working memory capacity than did fully congruent points ($r = .40$ vs. $r = .31$; trial type intercorrelation $r = .80$; $N = 223$; Steiger's $Z = 2.28$, $p = .023$). We also found that "stimulus congruent response incongruent" points correlated significantly more strongly with working memory capacity than did "stimulus incongruent response congruent" points ($r = .39$ vs. $r = .31$; trial type intercorrelation $r = .85$; $N = 223$; Steiger's $Z = 2.33$, $p = .020$).

For response times, the correlations across the different Squared task trial types ranged from $r = -.41$ to $r = -.48$ with attention control, from $r = -.50$ to $r = -.54$ with fluid intelligence, and from $r = -.29$ to $r = -.31$ with working memory capacity. Negative correlations indicate that participants who responded faster tended to have higher cognitive ability. The largest-magnitude difference in correlations was between attention control and fully incongruent response time versus "stimulus congruent response incongruent" response time ($r = -.41$ vs. $r = -.48$). Steiger's Z test indicated that this difference was statistically significant (trial type intercorrelation $r = .88$; $N = 223$; $Z = -2.39$, $p = .017$). The other differences between correlations were not statistically significant.

Study 2: Analysis of Trial Types

In the following section, we compare performance on the different trial types in the three Squared tasks and correlate performance on each trial type with other measures of cognitive ability. We report the results based on the first administration of each Squared task, noting that the results are broadly consistent when examining the other administrations of each task. As shown in Figure S2, participants earned the most points on fully congruent trials in all three Squared tasks. However, contrary to prediction, there was no pattern of performance differences across the other trial types that was consistent across the three Squared tasks. We ran a series of repeated measures ANOVAs to examine the effect of stimulus congruency, response option congruency, and their interaction on mean total points earned and mean response times (Table S3). We found significant main effects and interaction effects for everything except Simon Squared. For Simon Squared, we found non-significant main effects of stimulus congruency and response option congruency, but a significant interaction between stimulus congruency and response option congruency.

Table S3. Repeated-measures ANOVAs on the trial types in the Squared tasks (first in-lab administration, Study 2)

	Main effect of Stimulus Congruency	Main effect of Response Congruency	Interaction Effect
Stroop Squared Points	$F(1, 310) = 78.75$, $p < .001$, $\eta_p^2 = .203$	$F(1, 310) = 158.13$, $p < .001$, $\eta_p^2 = .338$	$F(1, 310) = 32.40$, $p < .001$, $\eta_p^2 = .095$
Flanker Squared Points	$F(1, 289) = 44.35$, $p < .001$, $\eta_p^2 = .133$	$F(1, 289) = 12.34$, $p = .001$, $\eta_p^2 = .041$	$F(1, 289) = 5.11$, $p = .024$, $\eta_p^2 = .017$
Simon Squared Points	$F(1, 309) = 0.44$, $p = .506$, $\eta_p^2 = .001$	$F(1, 309) = 6.20$, $p = .013$, $\eta_p^2 = .020$	$F(1, 309) = 24.74$, $p < .001$, $\eta_p^2 = .074$
Stroop Squared RT	$F(1, 310) = 37.17$, $p < .001$, $\eta_p^2 = .107$	$F(1, 310) = 421.15$, $p < .001$, $\eta_p^2 = .576$	$F(1, 310) = 42.41$, $p < .001$, $\eta_p^2 = .120$
Flanker Squared RT	$F(1, 289) = 159.98$, $p < .001$, $\eta_p^2 = .356$	$F(1, 289) = 101.37$, $p < .001$, $\eta_p^2 = .260$	$F(1, 289) = 10.52$, $p = .001$, $\eta_p^2 = .035$

NATURE AND MEASUREMENT OF ATTENTION CONTROL

Simon Squared	$F(1, 309) = 2.45,$	$F(1, 309) = 0.02,$	$F(1, 309) = 283.09,$
RT	$p = .119, \eta_p^2 = .008$	$p = .877, \eta_p^2 = .000$	$p < .001, \eta_p^2 = .478$

Note: η_p^2 = partial eta squared, a measure of the effect size.

NATURE AND MEASUREMENT OF ATTENTION CONTROL

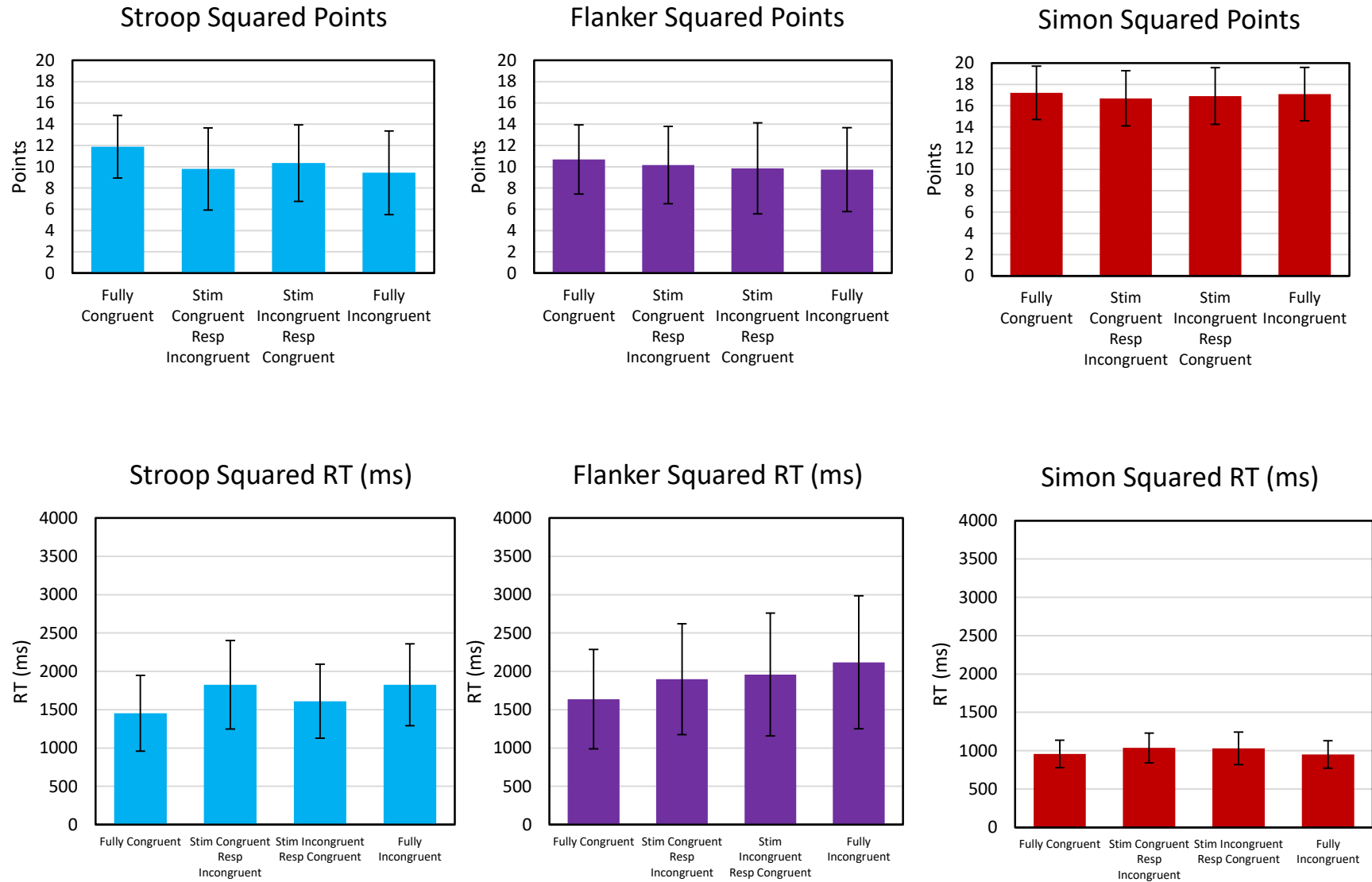


Figure S2. Performance on the four trial types in the Squared tasks (first in-lab administration). Error bars represent ± 1 SD (Study 2).

Next, we examined whether there were different patterns of correlations between performance on each of the trial types in the Squared tasks and the other cognitive ability factors (Table S4). We saved factor scores representing performance on the other attention control tests (i.e., antisaccade, SACT, and selective visual arrays), the fluid intelligence tests (i.e., Raven's matrices, letter sets, and number series), the working memory capacity tests (i.e., symmetry span, rotation span, and mental counters), and the processing speed tests (i.e., pattern comparison, letter comparison, and digit comparison) by performing four principal components analyses, one for each cognitive construct, and saving participants' scores on each factor. We used a similar approach for points earned and mean response times on each of the trial types in the Squared tasks. We then examined correlations between each of these markers of performance on the different trial types and the cognitive ability component scores.

Table S4. *Correlations between performance on each trial type in the Squared tasks and cognitive ability component scores (Study 2).*

Measure	Attention Control: Other	Fluid Intelligence	Working Memory Capacity	Processing Speed
Fully Congruent Points	.47	.53	.46	.57
Fully Incongruent Points	.53	.56	.47	.57
Stimulus Congruent Response Incongruent Points	.51	.52	.46	.52
Stimulus Incongruent Response Congruent Points	.49	.51	.44	.51
Fully Congruent RT	-.44	-.44	-.37	-.53
Fully Incongruent RT	-.45	-.51	-.46	-.53
Stimulus Congruent Response Incongruent RT	-.46	-.51	-.46	-.52
Stimulus Incongruent Response Congruent RT	-.45	-.52	-.45	-.51

Note: *N* ranges from 259 to 275. Each variable in this analysis is a component score extracted from a principal components analysis; one for each trial type or cognitive ability measure. Bold = statistically significant at $p < .05$.

For points earned, the correlations across the different Squared task trial types ranged from $r = .47$ to $r = .53$ with attention control, from $r = .51$ to $r = .56$ with fluid intelligence, from $r = .44$ to $r = .47$ with working memory capacity, and from $r = .51$ to $r = .57$ with processing speed.

Thus, all correlations with cognitive ability component scores were substantial and significant, and differences in the magnitude of the correlations across trial types were rather small and did not follow a consistent clear pattern. For response times, the correlations across the different Squared task trial types ranged from $r = -.44$ to $r = -.46$ with attention control, from $r = -.44$ to $r = -.52$ with fluid intelligence, from $r = -.37$ to $r = -.46$ with working memory capacity, and from $r = -.51$ to $r = -.53$ with processing speed. Negative correlations indicate that participants who responded faster tended to have higher cognitive ability composite scores. Once again, the trial-type correlations were substantial but did not appear to differ meaningfully across trial types. As different patterns of results emerged in Study 1 and Study 2, we urge caution in trying to interpret slight numerical differences in these correlations across trial types.

Study 2: Analyses Including Mental Counters

In Study 2, we reported the results of confirmatory factor analyses and structural equation models in which working memory capacity was identified using the two complex span tests, symmetry span and rotation span. Here, we report the results of these analyses when working memory capacity is identified using mental counters as well as the two complex span tests.

We examined correlations between attention control, working memory capacity, fluid intelligence, and processing speed at the latent level. We created two attention control latent factors—one for the three Squared tests and another for the other tests of attention control—and correlated each with the other cognitive ability factors. The purpose of this model was to examine how correlations between attention control and the other latent cognitive ability factors differed depending on how attention control was measured. The fit of the model was good ($\chi^2(80) = 179.63$ $p < .001$; CFI = .916, TLI = .890, RMSEA = .074, 90% CI [.059, .088], SRMR = .058); correlations are reported in Table S5.

Table S5. *Latent variable correlations between the Squared attention control factor, the other attention control factor, fluid intelligence, working memory capacity, and processing speed.*

Factor	1.	2.	3.	4.	5.
1. Squared Attention Control	---				
2. Other Attention Control	.76	---			
3. Fluid Intelligence	.73	.61	---		
4. Working Memory Capacity	.68	.88	.78	---	
5. Processing Speed	.77	.59	.68	.60	---

Accounting for the Positive Manifold

In our next analyses, we investigated whether attention control could account for the substantial positive correlations observed among the cognitive ability factors. We tested two models, one in which we defined attention control using the three Squared tasks and another in which we used the other attention control measures. In both models, attention control was

specified as a predictor of fluid intelligence, complex span working memory capacity, and processing speed. The residual variance in fluid intelligence, working memory capacity, and processing speed—that is, the variance that remained unaccounted for by attention control—was allowed to correlate. The purpose of these analyses was to determine the extent to which partialing out variance in attention control reduced the latent correlation between cognitive ability factors. If the residual correlations are reduced to a considerable degree, or to non-significance, this would provide evidence that attention control captures domain-general variance that is shared by a number of different cognitive constructs.

The model using the three Squared tasks is depicted in Figure S3 and fit the data well ($\chi^2(48) = 143.99, p < .001$; CFI = .910, TLI = .876, RMSEA = .090, 90% CI [.073, .107], SRMR = .063). Squared Attention control explained significant variance in each of the cognitive ability factors, with a standardized path of $\beta = .73$ ($p < .001$) to fluid intelligence, $\beta = .67$ ($p < .001$) to working memory capacity, and $\beta = .75$ ($p < .001$) to processing speed. Residual fluid intelligence correlated $r = .51, p < .001$ with residual working memory capacity (reduced significantly from $r = .78$, see Table S5; $\Delta\chi^2(1) = 7.46, p = .006$); residual fluid intelligence correlated $r = .26, p = .027$ with residual processing speed (reduced significantly from $r = .68$, see Table S5; $\Delta\chi^2(1) = 16.28, p < .001$); residual working memory capacity correlated non-significantly ($r = .14, p = .265$) with residual processing speed (reduced significantly from $r = .60$, see Table S5; $\Delta\chi^2(1) = 16.54, p < .001$). These results indicate that the Squared attention control factor partly explains the covariation between fluid intelligence and other cognitive abilities, and fully explains the covariation between working memory capacity and processing speed.

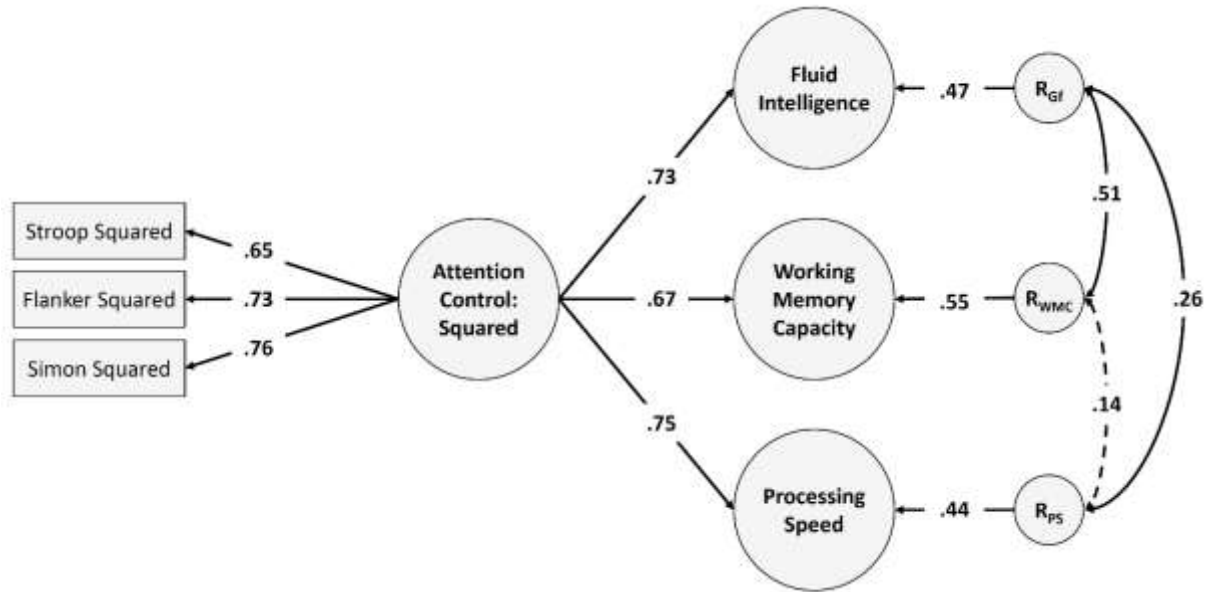


Figure S3. Structural equation model with a Squared attention control factor predicting fluid intelligence, complex span working memory capacity, and processing speed. The residual variance in each cognitive ability construct represents the variance in each construct after accounting for attention control. Indicators for fluid intelligence, working memory capacity, and processing speed are not depicted for visual clarity (Study 2).

The model using the three non-Squared tasks is depicted in Figure S4 and fit the data adequately ($\chi^2(48) = 121.23, p < .001$; CFI = .917, TLI = .886, RMSEA = .079, 90% CI [.062, .097], SRMR = .060). The non-Squared attention control factor explained significant variance in each of the cognitive ability factors, with a standardized path of $\beta = .62$ ($p < .001$) to fluid intelligence, $\beta = .89$ ($p < .001$) to working memory capacity, and $\beta = .60$ ($p < .001$) to processing speed. Residual fluid intelligence correlated $r = .72, p < .001$ with residual working memory capacity (reduced non-significantly from $r = .78$, see Table S5; $\Delta\chi^2(1) = 0.14, p = .71$); residual fluid intelligence correlated $r = .48, p < .001$ with residual processing speed (reduced significantly from $r = .68$, see Table S5; $\Delta\chi^2(1) = 6.11, p = .013$); residual working memory capacity correlated non-significantly ($r = .22, p = .236$) with residual processing speed (reduced

non-significantly from $r = .60$, see Table S5; $\Delta\chi^2(1) = 3.44, p = .064$).

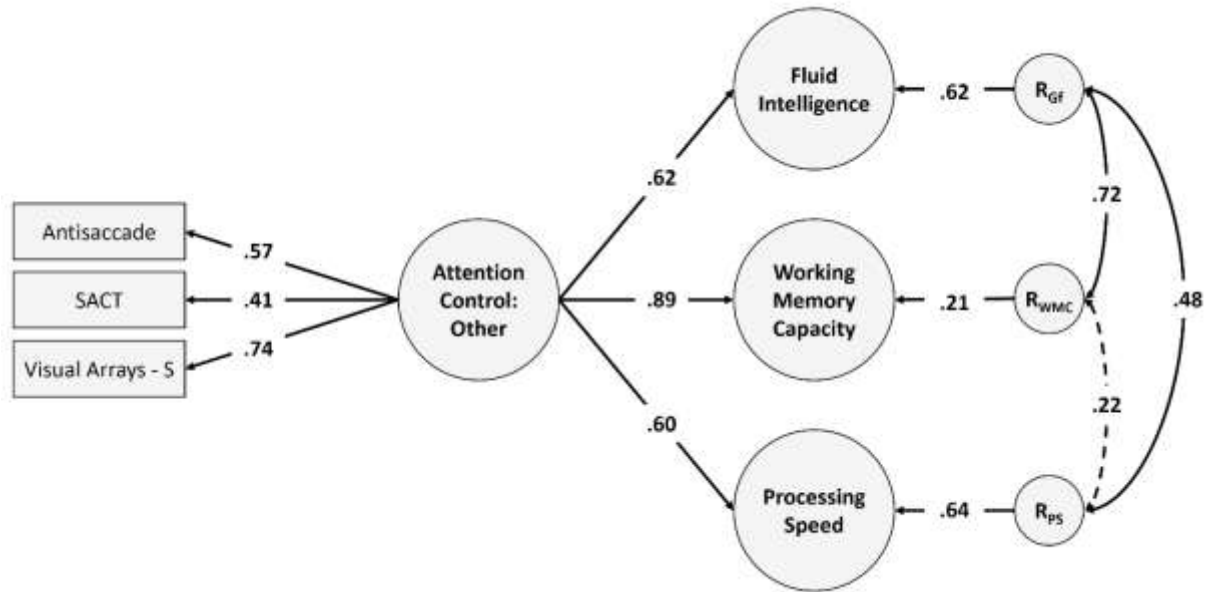


Figure S4. Structural equation model with a non-Squared attention control factor predicting fluid intelligence, working memory capacity, and processing speed. The residual variance in each cognitive ability construct represents the variance in each construct after accounting for attention control. Indicators for fluid intelligence, working memory capacity, and processing speed are not depicted for visual clarity.

Predicting Multitasking Ability

Last, we tested a model in which latent factors representing attention control, fluid intelligence, working memory capacity, and processing speed were all specified as predictors of multitasking ability. The predictor factors were allowed to correlate, allowing us to determine the relative contribution of each cognitive ability factor above and beyond the other factors. We tested two versions of this model, one with the Squared tasks and one with the other attention control tasks.

The model using the Squared tasks fit the data adequately ($\chi^2(94) = 205.55, p < .001$; CFI = .923, TLI = .901, RMSEA = .073, 90% CI [.059, .087], SRMR = .064). Fluid intelligence had the largest standardized path to multitasking ($\beta = .52$), followed by attention control ($\beta =$

.33), processing speed ($\beta = .22$), and working memory capacity ($\beta = .07$). Combined, the predictors accounted for 100% of the variance in multitasking.

Finally, we tested an identical model to the previous one except we used the three non-Squared tasks as indicators of attention control. The model fit the data adequately ($\chi^2 (94) = 186.20, p < .001$; CFI = .926, TLI = .906, RMSEA = .067, 90% CI [.052, .081], SRMR = .059). Fluid intelligence had the largest standardized path to multitasking ($\beta = .70$), followed by non-squared attention control ($\beta = .59$), processing speed ($\beta = .36$), and working memory capacity ($\beta = -.50$); the negative path from working memory capacity was unexpected. Combined, the predictors accounted for 100% of the variance in multitasking.