Nature and measurement of attention control

Article *in* Journal of Experimental Psychology General · April 2023

DOI: 10.1037/xge0001408

Alex Burgoyne
Georgia Institute of Technology

Jason S. Tsukahara
Georgia Institute of Technology

Cody Mashburn
Georgia Institute of Technology

Richard Pak
Clemson University

Journal of Experimental Psychology: General

Nature and Measurement of Attention Control

Alexander P. Burgoyne, Jason S. Tsukahara, Cody A. Mashburn, Richard Pak, and Randall W. Engle

Online First Publication, April 20, 2023. https://dx.doi.org/10.1037/xge0001408

CITATION

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023, April 20). Nature and Measurement of Attention Control. *Journal of Experimental Psychology: General*. Advance online publication. https://dx.doi.org/10.1037/xge0001408

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001408

Nature and Measurement of Attention Control

Alexander P. Burgoyne[1], Jason S. Tsukahara[1], Cody A. Mashburn[1], Richard Pak[2], and Randall W. Engle[1]
[1] School of Psychology, Georgia Institute of Technology
[2] Human Factors Institute, Clemson University

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 3 min 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 (avg. loading = .70), 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 measures of attention control. The tasks are freely available online: https://osf.io/7q598/.

Public Significance Statement
Reliably measuring individual differences in attention control has posed a challenge for the field. This paper reports the development and validation of three 90-s tests of attention control, dubbed the "Squared" tasks: Stroop Squared, Flanker Squared, and Simon Squared. The three Squared tasks demonstrated great internal consistency reliability and test–retest reliability, strong evidence for convergent validity with other measures of attention control, and explained a majority of the positive manifold and variance in multitasking ability. The three Squared tasks can be administered online via web browser, E-Prime, or as standalone programs for Mac and Windows (https://osf.io/7q598/). The three Squared tasks demonstrate that it is possible to reliably measure attention control at the observed and latent level by avoiding the use of response time difference scores. Furthermore, the measures reveal that individual differences in attention control can be represented as a unitary latent factor that is highly correlated with complex cognitive task performance.

Keywords: attention control, executive functions, measurement, multitasking

Supplemental materials: https://doi.org/10.1037/xge0001408.supp

Alexander P. Burgoyne https://orcid.org/0000-0002-2651-3782

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

A preprint of this manuscript was uploaded to PsyArXiv (https://psyarxiv.com/7y5fp/) on August 8, 2022. Study 2 data were included in a research poster presented at the 2022 Psychonomic Society Annual Meeting in Boston, MA on November 18, 2022. Data are openly available at https://osf.io/zkqbs/.

Alexander P. Burgoyne contributed toward conceptualization, methodology, software, formal analysis, investigation, resources, data curation, writing–original draft, writing–review and editing, visualization, and project administration. Jason S. Tsukahara contributed toward conceptualization, methodology, software, formal analysis, investigation, resources, data curation, writing–original draft, and writing–review and editing. Cody A. Mashburn contributed toward conceptualization, methodology, software, formal analysis, investigation, resources, data curation, writing–original draft, and writing–review and editing. Richard Pak contributed toward conceptualization, methodology, software, writing–original draft, and writing–review and editing. Randall W. Engle contributed toward conceptualization, methodology, investigation, resources, writing–original draft, writing–review and editing, supervision, project administration, and funding acquisition.

Correspondence concerning this article should be addressed to Alexander P. Burgoyne, School of Psychology, J.S. Coon Building, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332-0170, United States. Email: burgoyn4@gmail.com

Individual differences in the ability to control attention are correlated with a wide range of important outcomes, from cognitive task performance (Burgoyne, Mashburn, et al., 2023; Conway et al., 2002; Draheim et al., 2021, Draheim, Pak, et al., 2022; Engle et al., 1999; Martin, Mashburn, & Engle, 2020; 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., 2013; Schmeichel & Demaree, 2010; Zelazo & Cunningham, 2007). As such, considerable time and effort have 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 nonretrieval. Both functions require attention control, although they can also be modeled as separate but correlated latent factors (see Martin, Shipstead, et al., 2020).

Within the broader literature, attention control has 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 20 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 (i.e., perceptual 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., 2020).

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 & Sawi, 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.

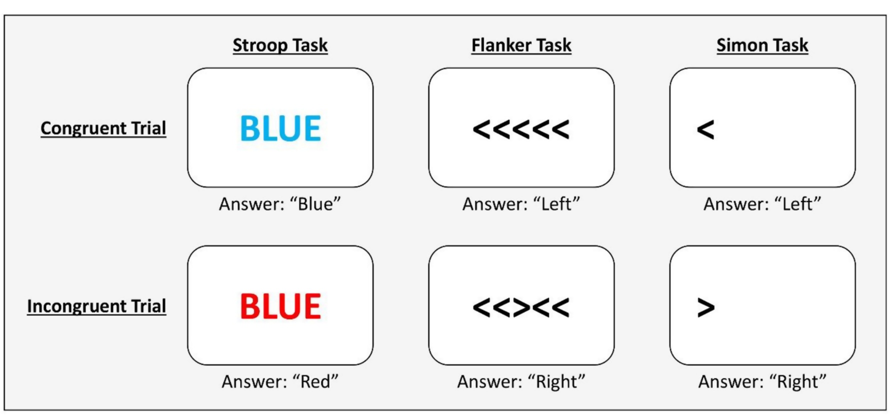
As is now well documented (Draheim et al., 2019; Hedge et al., 2018; Rouder & Haaf, 2019), part of the reliability problem can be attributed to psychometrically unsound tasks that use *response time difference scores* as the outcome measure, including "classic" experimental paradigms such as the Stroop (Stroop, 1935), Flanker (Eriksen & 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.

Furthermore, tasks that are well suited for experimental research are often poorly suited for individual differences research because they rely on an unreliable reaction-time difference score. 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 the conflict between the word's meaning and its color. By contrast, congruent trials require largely nonattentional processes because reading is 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,

Figure 1

Examples of Congruent and Incongruent Trials from the Classic Stroop, Flanker, and Simon Tasks



Note. 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. See the online article for the color version of this figure.

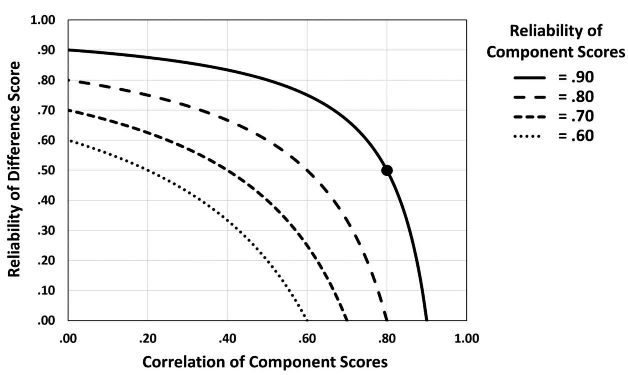
1991). Or, as Haaf and Rouder (2017) recently put it, "everybody Stroops" (p. 779). That said, psychometricians have cautioned against the use of difference scores in individual differences research for decades because of their unreliability at the level of the participant, and subsequently poor validity (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).

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), meaning that only 25% of its variance reflects the construct of interest! The consequence 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).

Figure 2

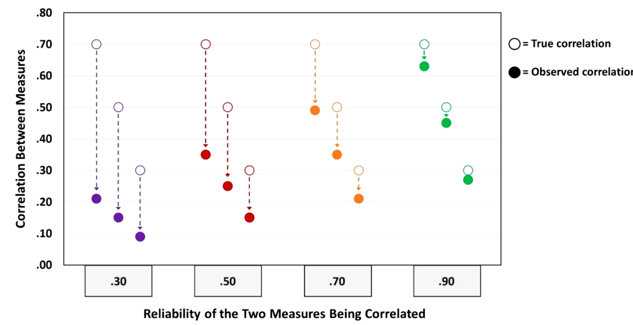
The Reliability of a Difference Score (Y-Axis) Decreases as the Correlation Between the Component Scores (i.e., Performance on Congruent Trials and Incongruent Trials) Increases (X-Axis)



Note. 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 "Reaction time in differential and developmental research: A review and commentary on the problems and alternatives," by C. Draheim, C. A. Mashburn, J. D. Martin, and R. W. Engle, 2019, *Psychological Bulletin*, *145*(5), pp. 508–535 (https://doi.org/10.1037/bul0000192). Copyright 2019 by the American Psychological Association.

Figure 3

The Attenuating Effect of Unreliability on the Observed Correlation Between Measures



Note. 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. See the online article for the color version of this figure.

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, 0.75), 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 6 months. Draheim et al. (2021) did not compute an internal consistency reliability coefficient (e.g., Cronbach's α 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 largely 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 2 to 12 s, 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 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). For 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 section.)

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 by creating a 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 incorporated 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, 1997), 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 precued 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 *nonselective 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 precue. Attentional filtering is crucial, because if a participant cannot selectively attend to the cued subset of items and block encoding and retention 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 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 takes less than 3 min 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 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 (Burgoyne et al., 2022; Conway et al., 2002; Draheim et al., 2021; Engle et al., 1999; Tsukahara et al., 2020; Unsworth et al., 2014), 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 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 toward 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 subtasks 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, Mashburn, and Engle (2020) 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, Mashburn, and Engle (2020) also explored whether processing speed accounted 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 s 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.

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 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.

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 five 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.

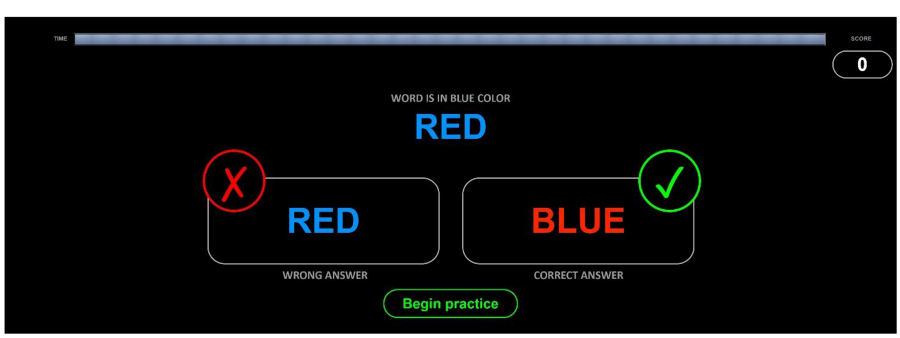
Simon Squared

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.

Trial Types in the Squared Tasks

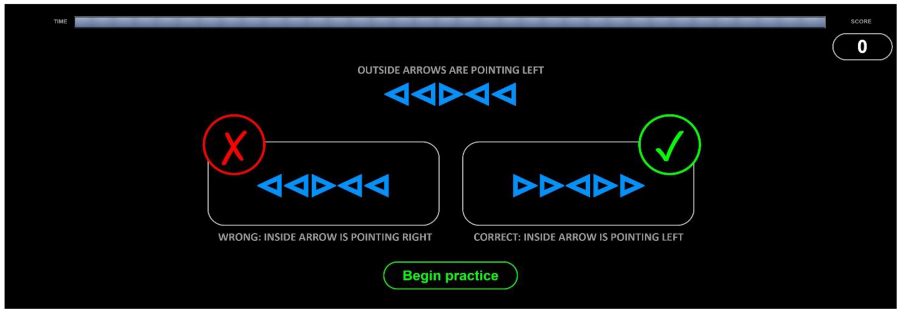
In each of the Squared tasks, there are four trial types that are sampled with equal probability (see Figure 7). Trial types are defined by whether the target stimulus and response options are "congruent," meaning the word's semantic meaning and display color match (e.g., "RED" in red color in Stroop Squared), or "incongruent," meaning the word's semantic meaning and display color do not

Figure 4
Stroop Squared



Note. 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). See the online article for the color version of this figure.

Figure 5
Flanker Squared



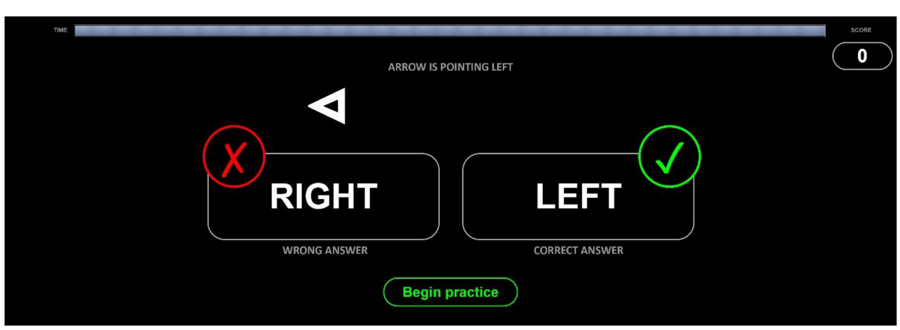
Note. The participant's task is to select the response option with a central arrow that points in the same direction as the flanking arrows in the target stimulus. In the above example, the target stimulus has flanking arrows pointing left, so the participant must select the response option which has a central arrow pointing left. See the online article for the color version of this figure.

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, and feedback is given on every trial), 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.

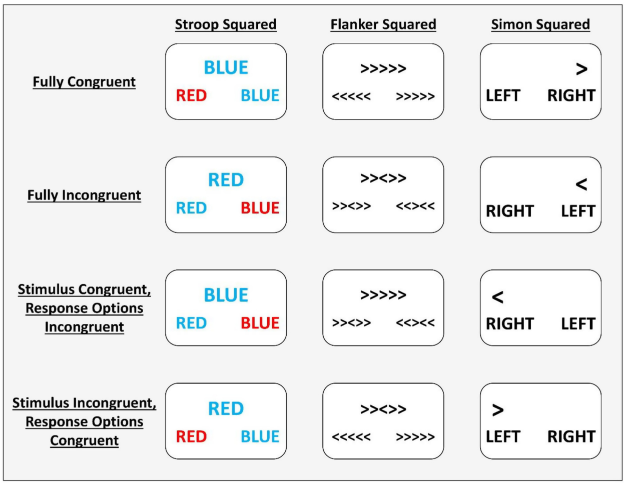
Figure 6
Simon Squared



Note. 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". See the online article for the color version of this figure.

Figure 7

Examples of the Four Trial Types in the Three Squared Tests of Attention Control



Note. 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. See the online article for the color version of this figure.

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 2–2.5 hr to complete. The tasks were programmed using E-Prime Go (Psychology Software Tools, Pittsburgh, PA, 2020) 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 and the description of the task that accompanies it. Participants completed a 30-s practice phase followed by a 90-s test phase. 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-s 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 s 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-s timer counted down, and then the test phase began. Participants were given 90 s 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-s 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 and the description of the task that accompanies it. Participants completed a 30 s practice phase followed by a 90 s test phase. 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 and the complete description of the task that accompanies it. Participants completed a 30-s practice phase followed by a 90-s test phase. 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 1,000 or 2,000 ms, an asterisk (*) flashed at 12.3° visual angle to the left or right of the central fixation for 100 ms. Afterward, the letter "Q" or "O" was presented on the opposite side at 12.3° visual angle of the central fixation for 100 ms, 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 750 ms, 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 trials) or all in the opposite direction (incongruent trials). 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 four 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 nonadaptive response deadline practice trials.

An adaptive staircase procedure was used to estimate the subject's response deadline that would converge around 60% accuracy. The adaptive procedure was based only on incongruent trials. On each incongruent trial, if an incorrect response was made *or* the response 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 response time was shorter than the response deadline, then the response deadline decreased (less time to respond) on the next trial. The initial value for the response deadline was 1.5 s. 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 240:80 ms, decreased to 120:40 ms after 17 incongruent trials, decreased to 60:20 ms after 33 incongruent trials, decreased to 30:10 ms after 49 incongruent trials, decreased to 15:5 ms after 65 incongruent trials, and finally settled at 9:3 ms 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; Draheim, Tsukahara, et al., 2022; 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 nontarget letters. Each trial started with a central black fixation for 1 s followed by a 750 ms 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 500 ms, and then was removed from the display during the wait time interval. The wait time lasted either 0 s or 2–12 s in 500 ms 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 250 ms. The target letter was identifiable as the central letter in slightly darker font color. The target and nontarget stimuli were B, P, or R's. The task had three 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 reattend 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 remind participants of the target spatial location during the wait interval.

Selective Visual Arrays (Adapted From Luck & Vogel, 1997). Participants were shown a fixation cross for 1,000 ms, followed by the word "RED" or "BLUE" that instructed them to pay attention to either the red or blue rectangles that would appear shortly. An array of red and blue rectangles arranged at different angle orientations (i.e., the "target array") appeared for 250 ms, which was followed by a blank screen lasting 900 ms. The display included three or five rectangles of each color. Afterward, an array appeared that included only the cued color of rectangles (i.e., the "probe array"), and a white dot was used to highlight one of the rectangles. The angle of this particular rectangle could be the same as it appeared in the target array, or different; both possibilities were equally likely. The participant's task was to use to determine whether the angle of the rectangle was the same or had changed, using the keyboard to respond. We used 48 trials for each set size, and computed capacity scores (k) for each set size using the single-probe correction (Cowan et al., 2005): set size × (hit rate + correction rejection rate − 1). The outcome measure was the mean k estimate across set sizes 3 and 5.

Fluid Intelligence

Raven's Advanced Progressive Matrices (Raven & Court, 1998). In Raven's matrices, participants were shown a grid of 3×3 line drawings patterns, with the pattern in the bottom-right corner missing. The participant's task was to select from six response options the pattern that best fit the array. We gave participants 10 min for 18 items from Raven's Advanced Progressive Matrices; the measure of performance was the number of items they correctly responded to.

Letter Sets (Ekstrom et al., 1976). In letter sets, participants were shown five sets of four letters and challenged to identify the set of letters that did not adhere to the same pattern as the others. We gave participants 10 min to complete 30 items; the measure of performance was the number of items they correctly responded to.

Number Series (Thurstone, 1938). In number series, we presented participants with a set of numbers that followed a pattern. They were shown four possible response options that could complete the pattern and needed to select the response option that best followed the pattern of the number series. We gave participants 5 min for 15 items; the measure of performance was the number of items they correctly responded to.

Working Memory Capacity

Advanced Symmetry Span (Unsworth et al., 2005). In symmetry span, participants must remember spatial locations while deciding whether patterns are symmetrical or not. On a given trial, the participant was shown a symmetrical or asymmetrical grid and needed to determine whether or not it was symmetrical. Next, they were shown a 4×4 grid of squares, and one of them was emphasized by a red color. Their goal was to memorize the location of the colored square. This symmetry/square interleaving pattern continued for 2–7 times (i.e., the set sizes used in the task). Afterward, the participant needed to report the location that the colored squares appeared in, in the order that they appeared. We gave participants 12 trials; two of each set size. We used the partial scoring method as the outcome measure of performance.

Advanced Rotation Span (Kane et al., 2004). In rotation span, participants remembered directional arrows while deciding whether a letter was in the proper orientation or mirror-imaged. On a given trial, the participant was shown a letter they would mentally rotate to determine its orientation (mirror-imaged or normal). Next, they were shown a single arrow that was either small or large and pointed in one of eight directions. This letter/arrow interleaving pattern continued 2–7 times (i.e., the set sizes used in the task). Afterward, the participant was asked to report the arrows in the order they appeared. We gave participants 12 trials; two of each set size. We used the partial scoring method as the outcome measure of performance.

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 5. Boxes would appear one at a time above or below the lines for 500–830 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 5 (e.g., five boxes appeared during the trial), 14 trials at set size 6, and 13 trials at set size 7, for a total of 32 trials. The measure of performance was the partial score, reflecting the number of correctly reported values.

Transparency and Openness

We report all data exclusions below. This study's design and its analysis were not pre-registered. Data for Study 1 and Study 2 are openly available on the Open Science Framework (https://osf.io/zkqbs). Data for Study 2 were collected as part of a larger project, the details of which are provided online (https://osf.io/qbwem).

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 subchance 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 200 ms, on the basis that these responses were too fast and likely represented misclicks. 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 *SD* 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 0.08 indicates a 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

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[a]: 19.2

Note. Demographic information was unavailable for some participants, lowering the effective *n* to 348. [a]Other includes, Hispanic or Latino, Native American, and "Other."

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.

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; avg. number of trials = 42), Flanker Squared (.94; avg. number of trials = 37), Simon Squared (.97; avg. 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).

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 $r = .46$).

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

Table 2
Descriptive Statistics for Study 1

Measure	*n*	*M*	*SD*	Skew	Kurtosis	Reliability	Time (min)
Stroop Squared	311	31.29	14.64	−0.45	−0.44	.93[b]	2
Flanker Squared	297	27.38	13.96	−0.14	−0.32	.94[b]	2
Simon Squared	321	57.38	15.26	−1.14	2.00	.97[b]	2
Antisaccade	306	0.79	0.13	−0.54	−0.60	.87[a]	8
FlankerDL	316	729.95	299.27	1.86	2.89	.89[a]	10
SACT	323	0.86	0.17	−1.50	1.32	.95[a]	18
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. Time = average administration time from starting to finishing the task. — = administration time was not measured for this task. [a]Cronbach's α. [b]Split-half reliability with Spearman–Brown correction.

Table 3
Task-Level Correlation Matrix for Study 1

Measure	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 $n = 205$).

with Factor 2: $r = .71$; Factor 1 with Factor 3: $r = .51$, Factor 2 with Factor 3: $r = .51$).

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 0.68. The other attention control measures had slightly lower loadings: antisaccade (.57), SACT (.46), and selective visual arrays (.51). The exception was FlankerDL, which had a nonsignificant 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.

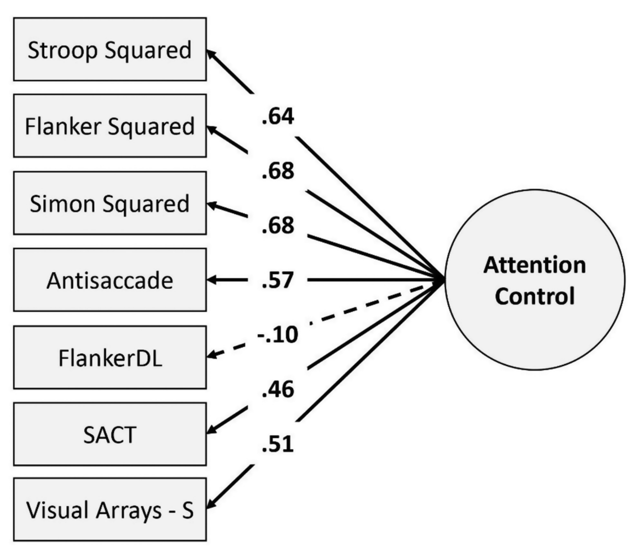
Table 4
Exploratory Factor Analysis for Study 1

Measure	Factor 1 (AC)	Factor 2 (Gf)	Factor 3 (WMC)
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. AC = attention control, Gf = fluid intelligence, WMC = working memory capacity.

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 fits the data well; $\chi^2(8) = 2.76$, $p = .949$; CFI = 1.00, TLI = 1.04, RMSEA = 0.000, 90% CI [0.000, 0.006], SRMR = 0.017. 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

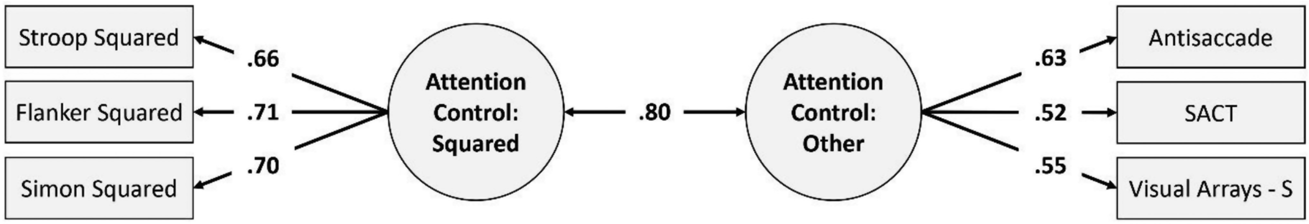
Figure 8
Latent Variable Model With All Attention Control Measures Loading on a Common Factor (Study 1)



Note. $\chi^2(14) = 14.70$, $p = .42$; CFI = 0.998, TLI = 0.997, RMSEA = 0.012, 90% CI [0.000, 0.068], SRMR = 0.036. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root mean square residual.

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)



Note. $\chi^2(8) = 2.76$, $p = .949$; CFI = 1.00, TLI = 1.04, RMSEA = 0.000, 90% CI [0.000, 0.006], SRMR = 0.017.

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

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 = 0.903, TLI = 0.866, RMSEA = 0.085, 90% CI [0.066, 0.104], SRMR = 0.072.

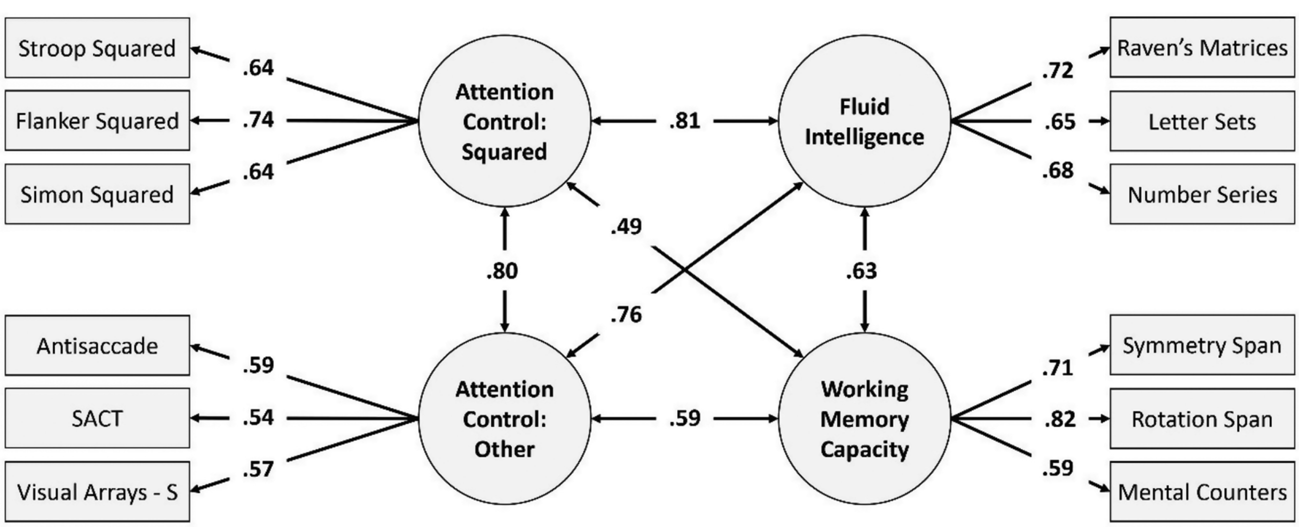
Structural Equation Modeling

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

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)



Note. $\chi^2(48) = 120.88$, $p < .001$; CFI = 0.903, TLI = 0.866, RMSEA = 0.085, 90% CI [0.066, 0.104], SRMR = 0.072.

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 = 0.924, TLI = 0.887, RMSEA = 0.096, 90% CI [0.073, 0.120], SRMR = 0.076. 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 Figures 10 and 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.

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 = 0.910, TLI = 0.865, RMSEA = 0.095, 90% CI [0.071, 0.119], SRMR = 0.073. 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 Figures 10 and 12). 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.

Analysis of Trial Types

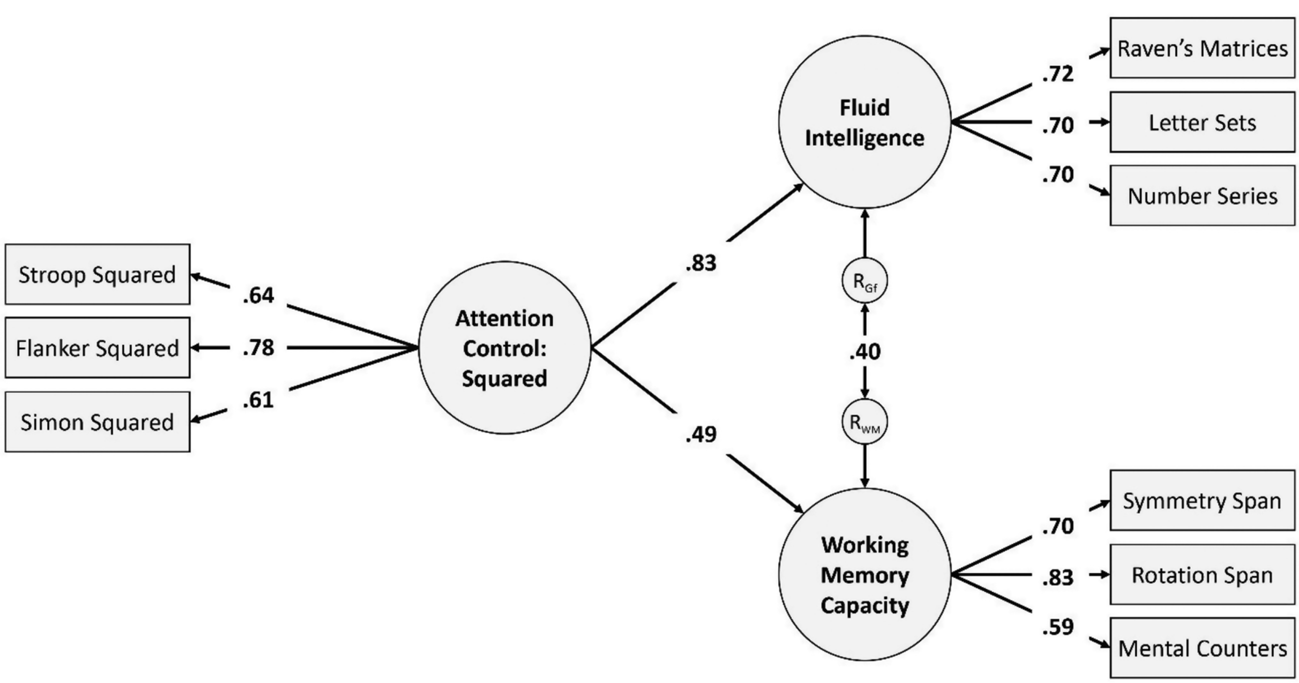
See the online supplemental materials.

Discussion of Study 1

Study 1 established that the 3-min "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 nonsignificant 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

Figure 11

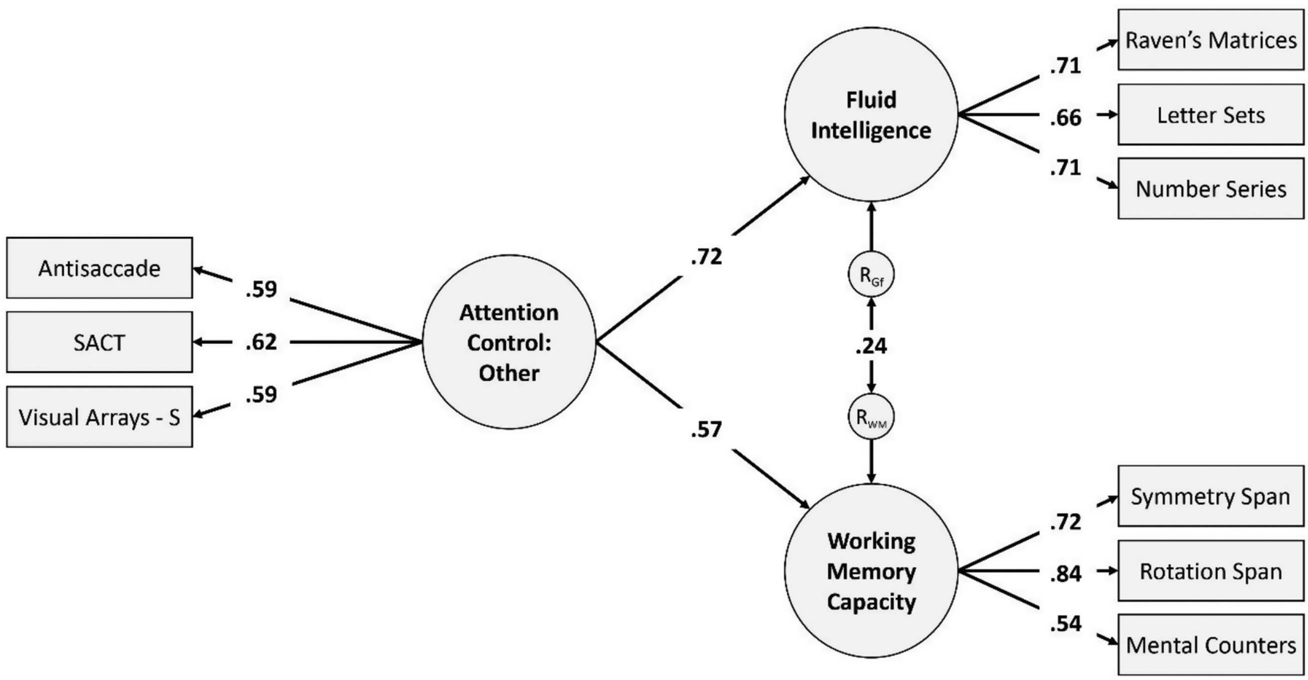
Structural Equation Model With a Latent Factor for the Three Squared Tests of Attention Control Predicting Fluid Intelligence and Working Memory Capacity



Note. $\chi^2(24) = 79.16$, $p < .001$; CFI = 0.924, TLI = 0.887, RMSEA = 0.096, 90% CI [0.073, 0.120], SRMR = 0.076). 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).

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



Note. $\chi^2(24) = 77.13$, $p < .001$; CFI = 0.910, TLI = 0.865, RMSEA = 0.095, 90% CI [0.071, 0.119], SRMR = 0.073. 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).

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 4). 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 the online supplemental materials).

Study 2: In-Laboratory Study

There are a few limitations of Study 1 that motivated 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 Burgoyne, Hambrick, & Altmann, 2021; Martin, Mashburn, & Engle, 2020). 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, Georgia, USA. All participants were required to be native English speakers and 18–35 years of age. 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 hr 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 hr 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 phones 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 seven 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 were shown 3, 6, or 9 numbers that appeared on the left and right side of a horizontal line drawn between them. The participant's task was to determine whether the strings of digits were identical or different. They responded using the mouse. Participants were given two blocks of 30 s of trials and attempted to answer as many items correctly as possible. Participants earned one point for each correct response and lost one point for each incorrect response; the measure of performance was the number of points earned at the conclusion of the task.

Letter String Comparison (Redick et al., 2012; Salthouse & Babcock, 1991). This task was almost identical to the digit string comparison task, however, instead of digits, the participant made comparisons about strings of letters.

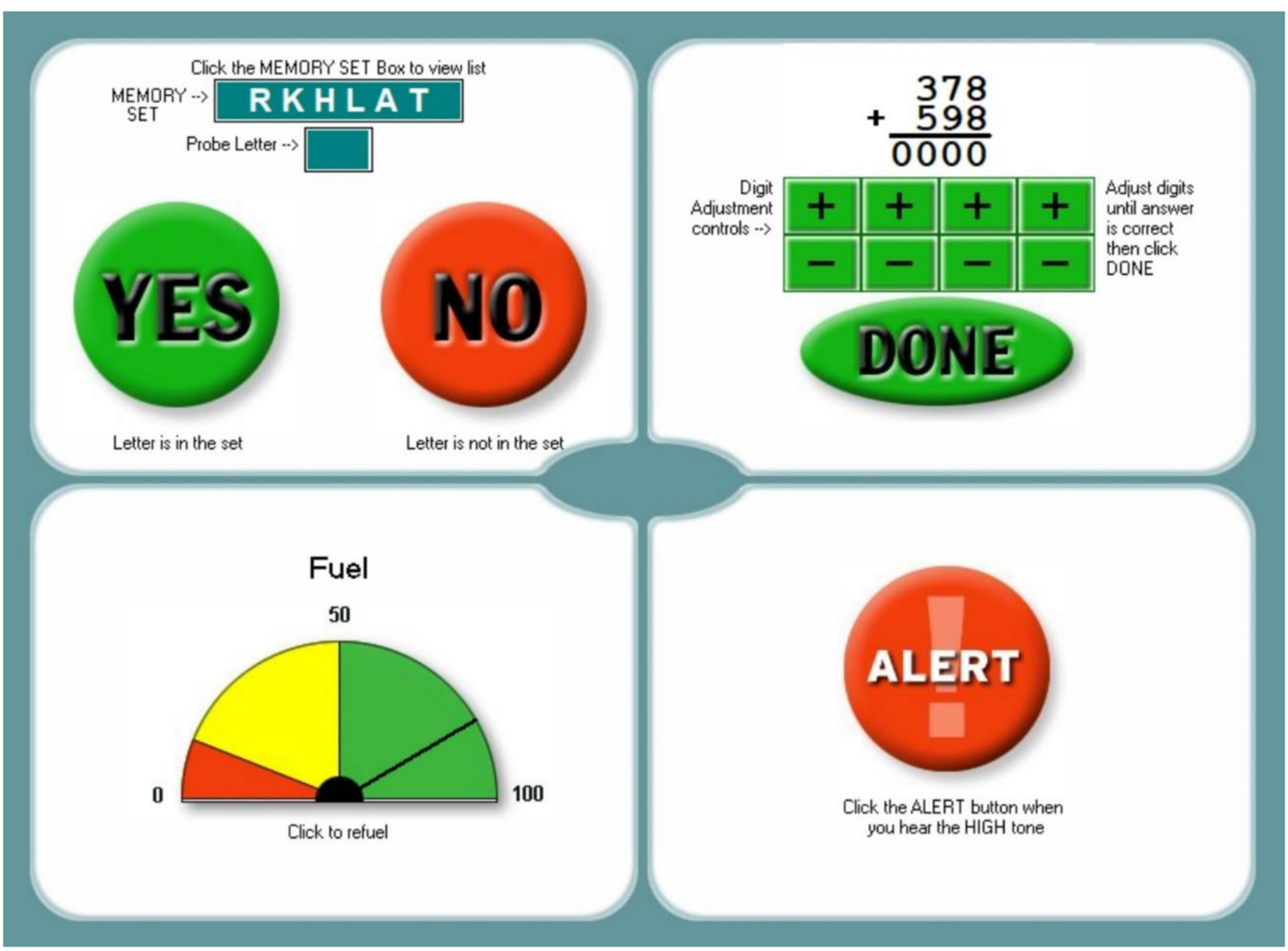
Pattern Comparison (Salthouse & Babcock, 1991). The participant was shown two symbols that appeared on either side of a horizontal line and indicated whether they were the same or different. Participants were given two blocks of 30 s of trials and attempted to answer as many items correctly as possible. Participants earned one point for each correct response and lost one point for each incorrect response; the measure of performance was the number of points earned at the conclusion of the task.

Multitasking Paradigms

Synthetic Work for Windows (SynWin; Elsmore, 1994; Figure 13). In SynWin, participants must manage four subtasks to earn as many points as possible. The subtasks included memory search, mathematics, and visual and auditory monitoring. Task details are presented in Martin, Mashburn, and Engle (2020). The

Figure 13

Synthetic Work (SynWin)



Note. The four subtasks are: Memory Search (top-left); Math (top-right); Visual Monitoring (bottom-left); and Auditory Monitoring (bottom-right). See the online article for the color version of this figure.

outcome measure was the average score across three 5-min test blocks.

Foster Multitask (Martin et al., 2020; Figure 14). The four subtasks included mathematics, word recall, and two visual monitoring subtasks. The outcome measure was the average score across three 5-min test blocks. Task details are presented in Martin, Mashburn, and Engle (2020).

Control Tower (Redick et al., 2016; Figure 15). Participants were given a primary task and multiple distractor tasks to complete over one 10-min block. The primary task entailed a symbol substitution task involving numbers, letters, and symbols. The distractor tasks included radar monitoring, problem solving, color identification, and clearing virtual airplanes for landing. The primary score was the number of symbol substitutions that were accurately performed, whereas the distractor score was the total number of correct responses given to the distractor tasks. Further details are provided in Martin, Mashburn, and Engle (2020).

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 or multitasking ability. For FlankerDL, we set trial-level performance to missing if the response time on that trial was less than 200 ms, on the basis that these responses were too fast and likely represented misclicks. 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 *SD* 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 one 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

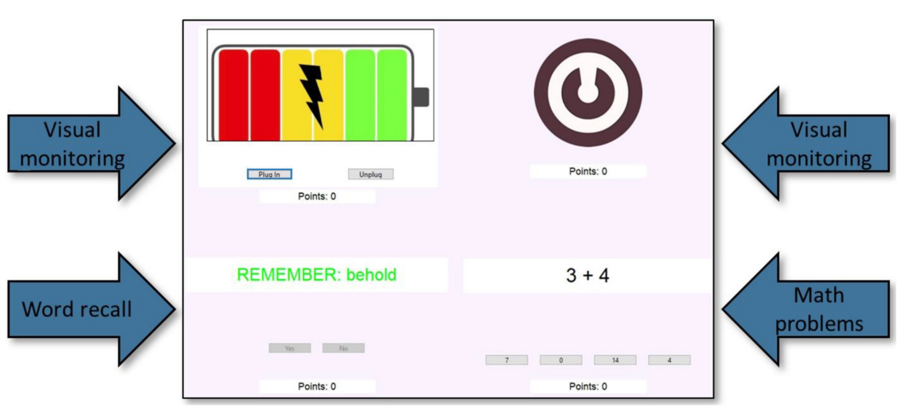
Demographic information is reported in Table 5. The participants' average age was 22 ($SD = 4$) years old and a majority were female (58.9%). Our in-lab sample was slightly older than the online sample from Study 1, which had a mean age of 27—the difference was statistically significant, $t(660) = 14.80$, $p < .001$. 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.

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.

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

Figure 14

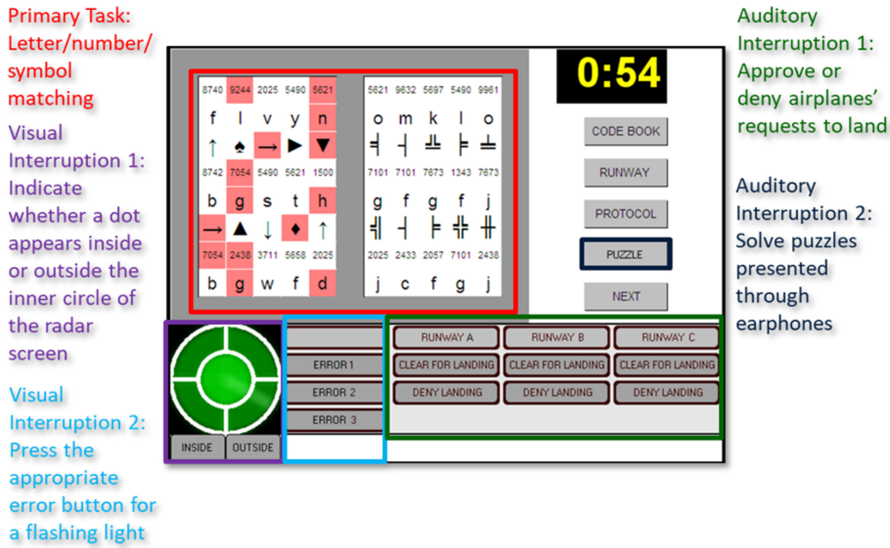
A Labeled Snapshot of the Foster Multitask Interface



Note. See the online article for the color version of this figure.

Figure 15

Labeled Snapshot of the Control Tower Interface



Note. See the online article for the color version of this figure.

that the third administration of the test occurred outside of the laboratory on the participants' personal computers approximately 2 months later: 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 test–retest–retest reliability, even across testing environments. 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.

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, $d = -0.35$, $t(54) = 2.47$, $p = .017$, Simon Squared, $d = -0.25$, $t(55) = 1.83$, $p = .072$. Thus, the Squared tasks demonstrated surprising resistance to practice effects; within-participant changes in performance were very small and generally nonsignificant across testing administrations.

We conducted a repeated-measures analysis of variance 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 nonsignificant, $F(2, 55) = 1.54$, $p = .223$, $\eta_p^2 = 0.053$; the effect of test administration on Flanker Squared was marginally significant, such that scores decreased over time, $F(2, 53) = 3.09$, $p = .054$, $\eta_p^2 = 0.104$; the effect of test administration on Simon Squared was marginally significant, such that scores decreased over time, $F(2, 54) = 3.16$, $p = .050$, $\eta_p^2 = 0.105$.

Table 5

Demographic Information for Study 2

Demographic	Statistic
Age (years)	Mean = 21.95
	SD = 4.09
	Range = 18–35
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[a]: 16.9

Note. $n = 314$. [a]Other includes, Hispanic or Latino, Native American, and Other.

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,

Table 6
Descriptive Statistics for Study 2

Measure	*n*	*M*	*SD*	Skew	Kurtosis	Reliability	Time (min)
Stroop Squared 1	311	41.44	12.76	−0.11	0.17	.94[b]	2
Stroop Squared 2	311	39.95	13.18	−0.29	0.52	.94[b]	2
Stroop Squared 3	63	41.06	16.76	−0.98	0.48	.95[b]	2
Flanker Squared 1	290	40.43	14.17	0.10	−0.20	.97[b]	2
Flanker Squared 2	290	39.47	14.13	0.14	0.26	.96[b]	2
Flanker Squared 3	63	31.83	15.80	−0.15	−0.59	.94[b]	2
Simon Squared 1	310	67.87	9.34	−0.20	0.24	.94[b]	2
Simon Squared 2	310	67.32	9.03	−0.22	−0.07	.95[b]	2
Simon Squared 3	61	63.85	12.91	−0.85	1.31	.96[b]	2
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
SACT	307	0.89	0.10	−1.11	0.71	.87[a]	17
Visual arrays	316	2.47	0.70	−0.51	0.07	.91[b]	12
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]	—
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	3,243.83	568.57	−0.53	1.14	.90[a]	—
Foster multitask	302	96,011.66	26,343.41	−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. — = 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. [a]Cronbach's α. [b]Split-half reliability with Spearman–Brown correction.

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

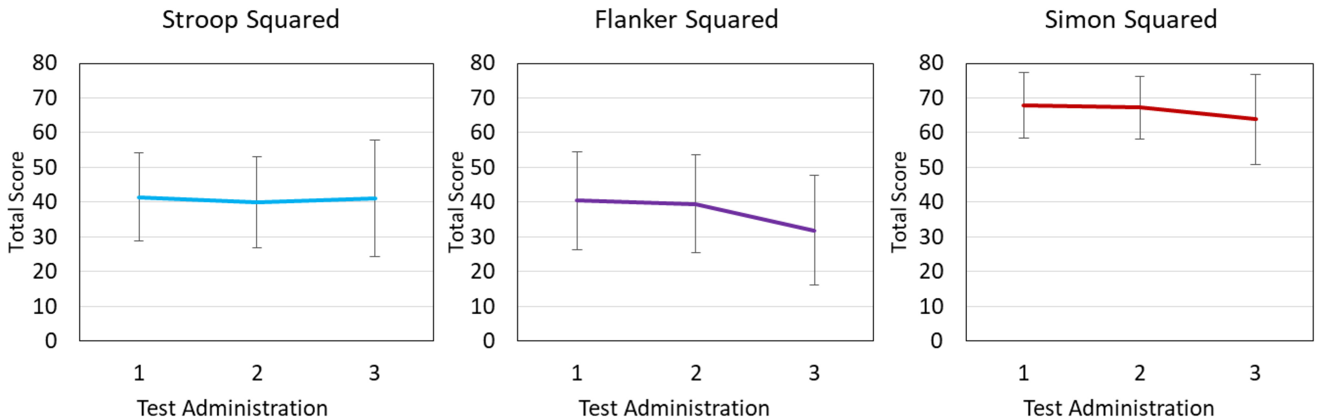
Table 7
Correlations Between Each Administrations of the Squared Tasks from Study 2

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., Stroop^2 = Stroop Squared). The number following each task name indicates the test administration number. *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). Bold = Statistically significant ($p < .05$).

Figure 16

Scores on Each of the Three Squared Tasks Across the Three Test Administrations



Note. Error bars represent ±1 *SD* around the mean (Study 2). See the online article for the color version of this figure.

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.

Confirmatory Factor Analyses

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 multiple attempts on the other tasks, so using participants' first attempt on the Squared tasks would make factor loadings more interpretable. We conducted a series of confirmatory factor analyses, first testing 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 = 0.930, TLI = 0.884, RMSEA = 0.105 90% CI [0.069, 0.143], SRMR = 0.055. 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).

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 = 0.961, TLI = 0.927, RMSEA = 0.083, 90% CI [0.043, 0.125], SRMR = 0.042. 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.

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 working memory capacity factor was defined using the two complex span measures. (Results including mental counters as an indicator of working memory capacity are provided in the online supplemental materials). The fit of the model was good; $\chi^2(67) = 137.54$, $p < .001$; CFI = 0.934, TLI = 0.910, RMSEA = 0.067, 90% CI [0.051, 0.083], SRMR = 0.054.

As shown in Table 9, 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, $\Delta\chi^2(1) =$

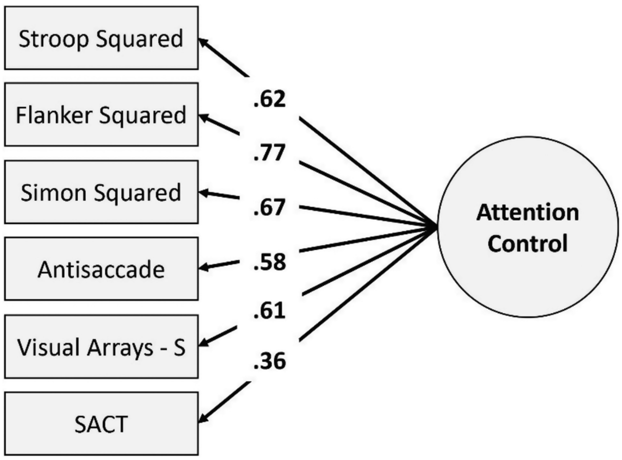
Table 8

Task-Level Correlation Matrix for Study 2

Measure	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. 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 (ns for those tasks ranged from 53 to 60). Control Tower (P) = Primary score; (D) = Distractor score. Boldface indicates $p < .05$.

Figure 17
Latent Variable Model With All Attention Control Measures Loading on a Common Factor (Study 2)



Note. $\chi^2(9) = 34.65$, $p < .001$; CFI = 0.930, TLI = 0.884, RMSEA = 0.105 90% CI [0.069, 0.143], SRMR = 0.055.

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.

Accounting for the Positive Manifold

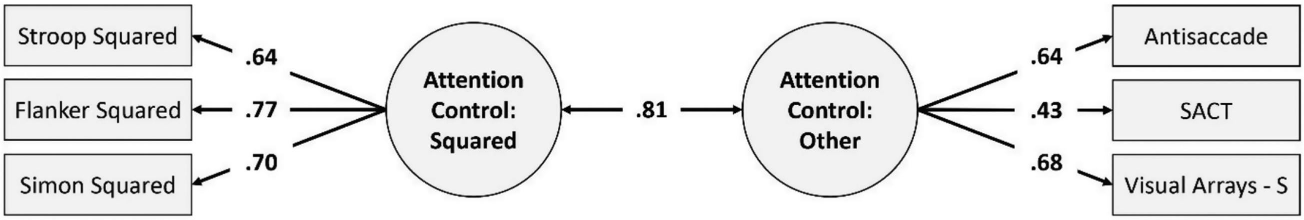
In our next analyses, we investigated whether attention control accounted for the substantial positive correlations observed among the cognitive ability factors. We tested two models, one in which we defined attention control using the three Squared tasks and another in which we used the other attention 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 nonsignificance, 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 = 0.931, TLI = 0.900, RMSEA = 0.083, 90% CI [0.064, 0.102], 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 significantly lower after accounting for attention control. Residual fluid intelligence correlated $r = .28$, $p = .008$ with residual working memory capacity (reduced significantly from $r = .53$, see Table 9); $\Delta\chi^2(1) = 6.61$, $p = .010$; residual fluid intelligence correlated $r = .29$, $p = .011$ with residual processing speed (reduced significantly from $r = .68$, see Table 9); $\Delta\chi^2(1) = 15.62$, $p < .001$; residual working memory capacity correlated nonsignificantly ($r = .13$ $p = .260$) with residual processing speed (reduced significantly from $r = .45$, see Table 9) $\Delta\chi^2(1) = 8.46$, $p = .004$. 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.

The model using the three other attention control tasks is depicted in Figure 20; $\chi^2(38) = 82.01$, $p < .001$; CFI = 0.943, TLI = 0.917, RMSEA = 0.068, 90% CI [0.047, 0.088], SRMR = 0.054. 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 $r = .53$, see Table 9); $\Delta\chi^2(1) = 7.36$, $p = .007$; residual fluid intelligence correlated $r = .47$, $p < .001$ with residual processing speed (reduced significantly from $r = .68$, see Table 9); $\Delta\chi^2(1) = 7.05$, $p = .008$; residual working memory capacity correlated nonsignificantly ($r = .16$, $p = .176$) with residual processing speed (reduced significantly from $r = .45$, see Table 9); $\Delta\chi^2(1) = 7.54$, $p = .006$. In other words, the non-Squared attention control factor accounted for a significant portion of the positive manifold.

Finally, we investigated whether processing speed could account for the positive correlations observed among the cognitive ability

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)



Note. $\chi^2(8) = 22.37$, $p = .004$; CFI = 0.961, TLI = 0.927, RMSEA = 0.083, 90% CI [0.043, 0.125], SRMR = 0.042.

Table 9

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	.77	—			
3. Fluid intelligence	.71	.61	—		
4. Working memory capacity	.52	.61	.53	—	
5. Processing speed	.76	.60	.68	.45	—

Note. All correlations were statistically significant at $p < .05$.

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 = 0.918, TLI = 0.895, RMSEA = 0.072, 90% CI [0.057, 0.087], SRMR = 0.058. 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 $r = .45$, $p < .001$ 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 nonsignificance.

Predicting Multitasking Ability

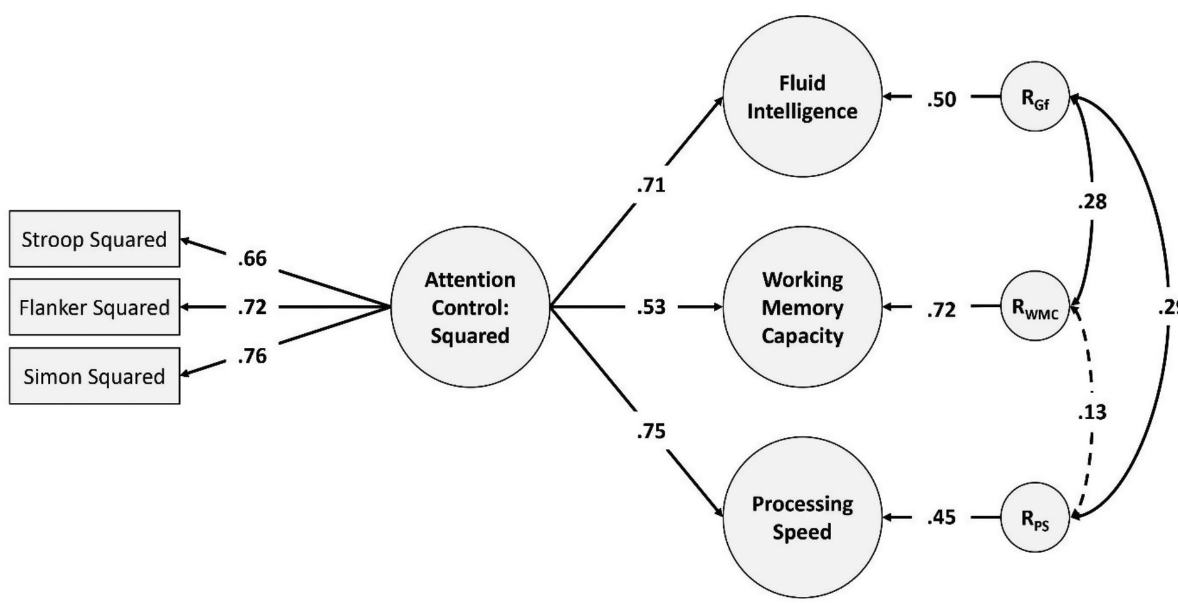
In this final section of analyses, we examine the relative contributions of different cognitive ability factors to multitasking ability. Our multitasking factor included four observed measures from three paradigms (SynWin, the Foster Multitask, and Control Tower: Primary and Distractor scores). These multitasking paradigms challenge participants to manage multiple information processing demands simultaneously (or concurrently), including elements of visual and auditory processing, arithmetic, memory, symbol substitution, 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 and fits the data well; $\chi^2(13) = 27.95$, $p = .009$; CFI = 0.975, TLI = 0.959, RMSEA = 0.067, 90% CI [0.032, 0.102], SRMR = 0.037.

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 = 0.990, TLI = 0.983, RMSEA = 0.036, 90% CI [0.000, 0.076], SRMR = 0.037.

Figure 19

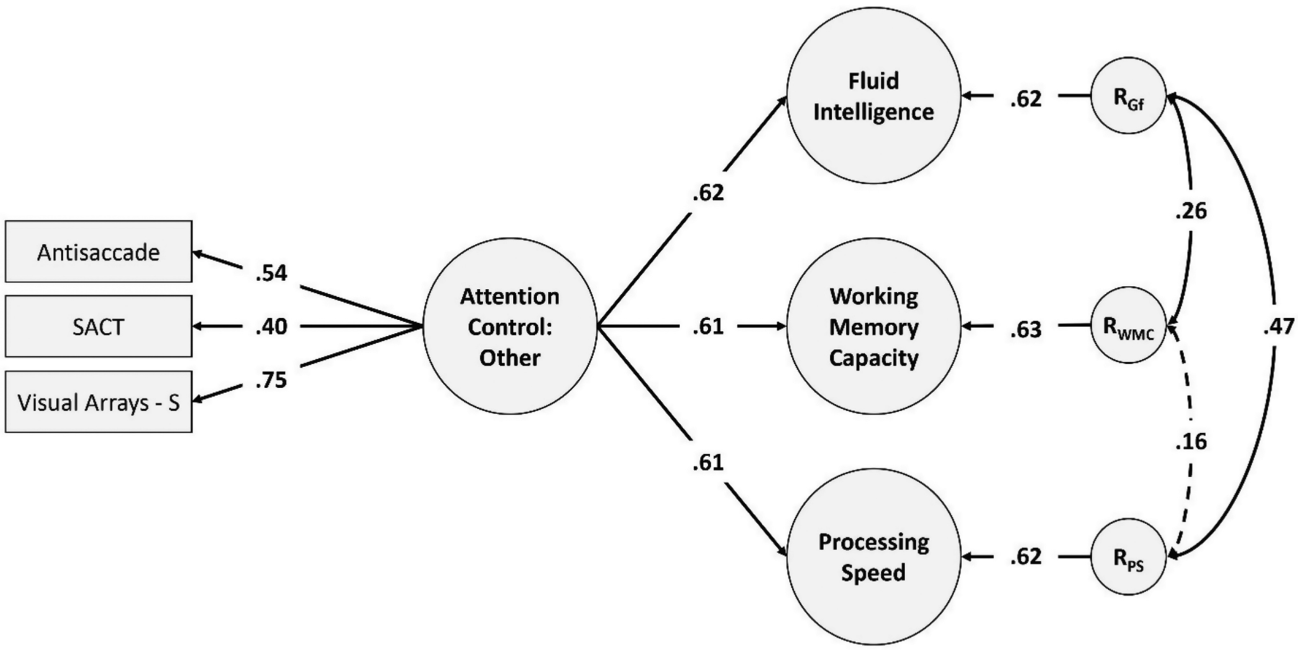
Structural Equation Model With a Squared Attention Control Factor Predicting Fluid Intelligence, Complex Span Working Memory Capacity, and Processing Speed



Note. 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). $\chi^2(38) = 103.67$, $p < .001$; CFI = 0.931, TLI = 0.900, RMSEA = 0.083, 90% CI [0.064, 0.102], SRMR = 0.057.

Figure 20

Structural Equation Model With a Non-Squared Attention Control Factor Predicting Fluid Intelligence, Working Memory Capacity, and Processing Speed



Note. 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. $\chi^2(38) = 82.01$, $p < .001$; CFI = 0.943, TLI = 0.917, RMSEA = 0.068, 90% CI [0.047, 0.088], SRMR = 0.054.

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 = 0.950, TLI = 0.930, RMSEA = .069, 90% CI [0.047, 0.092], SRMR = 0.049. 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 ($\beta = .23$). That said, setting the predictive paths equal to the same constant (i.e., "x") did not significantly worsen model fit, $\Delta\chi^2(1) = 2.95$, $p = .086$. Combined, the two attention control factors accounted for 76.8% of the variance in 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; $\chi^2(80) = 163.20$, $p < .001$; CFI = 0.938, TLI = 0.919, RMSEA = 0.068, 90% CI [0.053, 0.082], SRMR = 0.057. Fluid intelligence had the largest standardized path to multitasking ($\beta = .57$, $p < .001$), followed by attention control ($\beta = .32$, $p < .001$), processing speed ($\beta = .20$, $p = .029$), and working memory capacity ($\beta = .07$, $p = .26$). 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.

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 = 0.944, TLI = 0.927, RMSEA = 0.060, 90% CI [0.044, 0.075], SRMR = 0.055. Fluid intelligence had the largest standardized path to multitasking ($\beta = .56$, $p < .001$), followed by processing speed ($\beta = .36$, $p < .001$), non-Squared attention control ($\beta = .23$, $p = .034$), and working memory capacity ($\beta = -.01$, $p = .924$). Combined, the predictors accounted for 98.7% of the variance in multitasking.

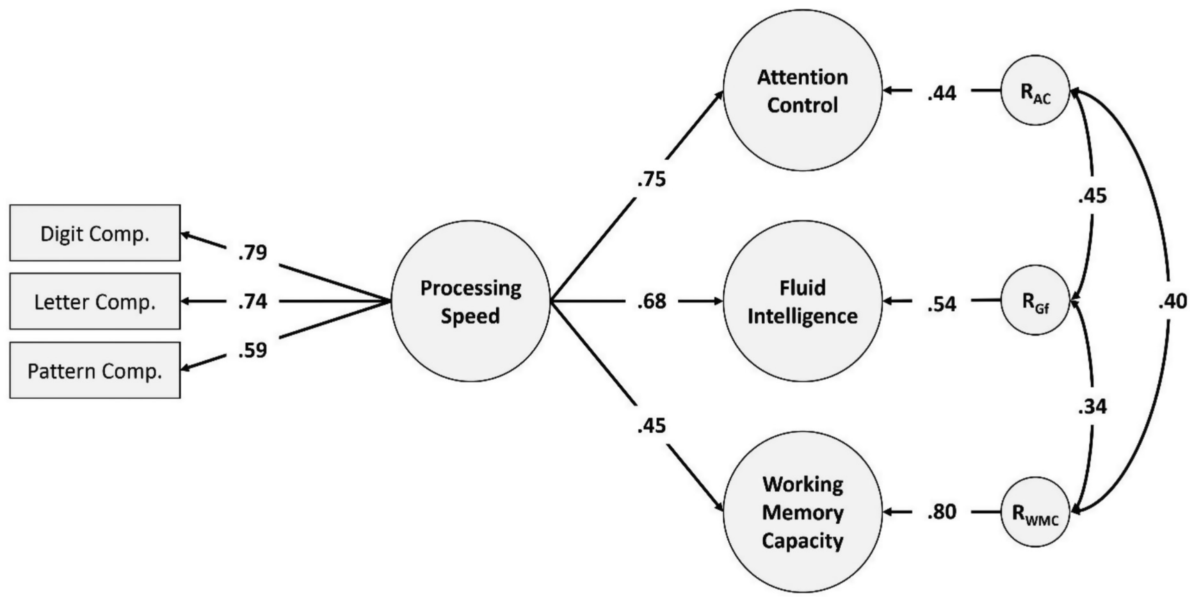
Analysis of Trial Types

See the online 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.

Figure 21
Structural Equation Model With a Processing Speed Factor Predicting Attention Control, Fluid Intelligence, and Working Memory Capacity



Note. 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. $\chi^2(71) = 158.03$, $p < .001$; CFI = 0.918, TLI = 0.895, RMSEA = 0.072, 90% CI [0.057, 0.087], SRMR = 0.058.

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).

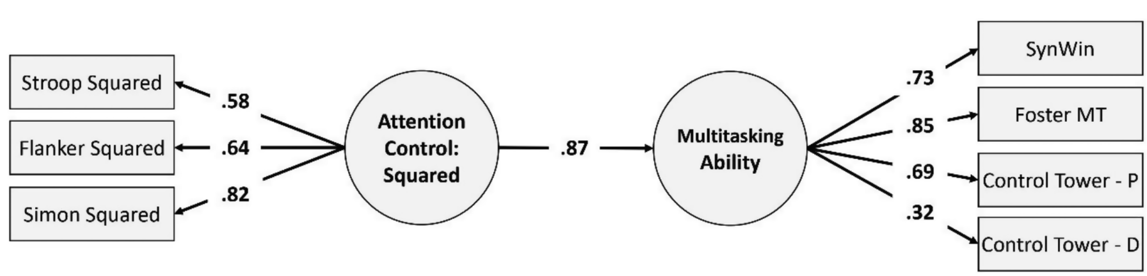
Internal Consistency and Test–Retest Reliability

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' personal 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

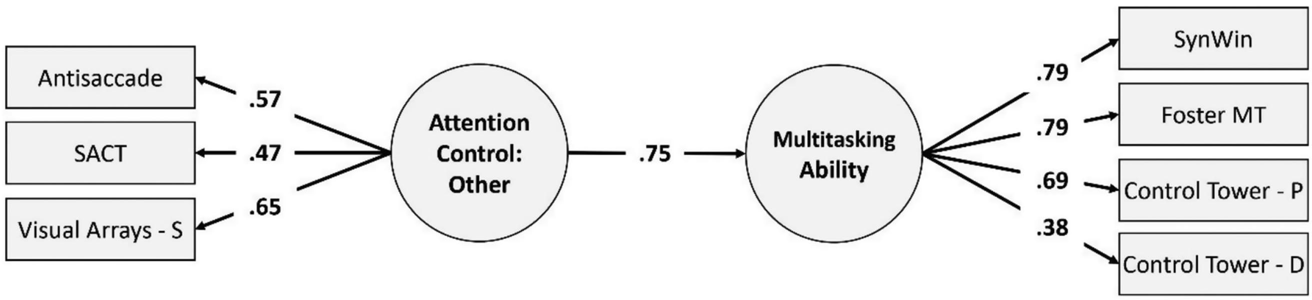
Figure 22
Structural Equation Model With a Squared Attention Control Factor Predicting Multitasking Ability



Note. $\chi^2(13) = 27.95$, $p = .009$; CFI = 0.975, TLI = 0.959, RMSEA = 0.067, 90% CI [0.032, 0.102], SRMR = 0.037.

Figure 23

Structural Equation Model With a Non-Squared Attention Control Factor Predicting Multitasking Ability



Note. $\chi^2(13) = 17.32$, $p = .185$; CFI = 0.990, TLI = 0.983, RMSEA = 0.036, 90% CI [0.000, 0.076], SRMR = 0.037.

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.

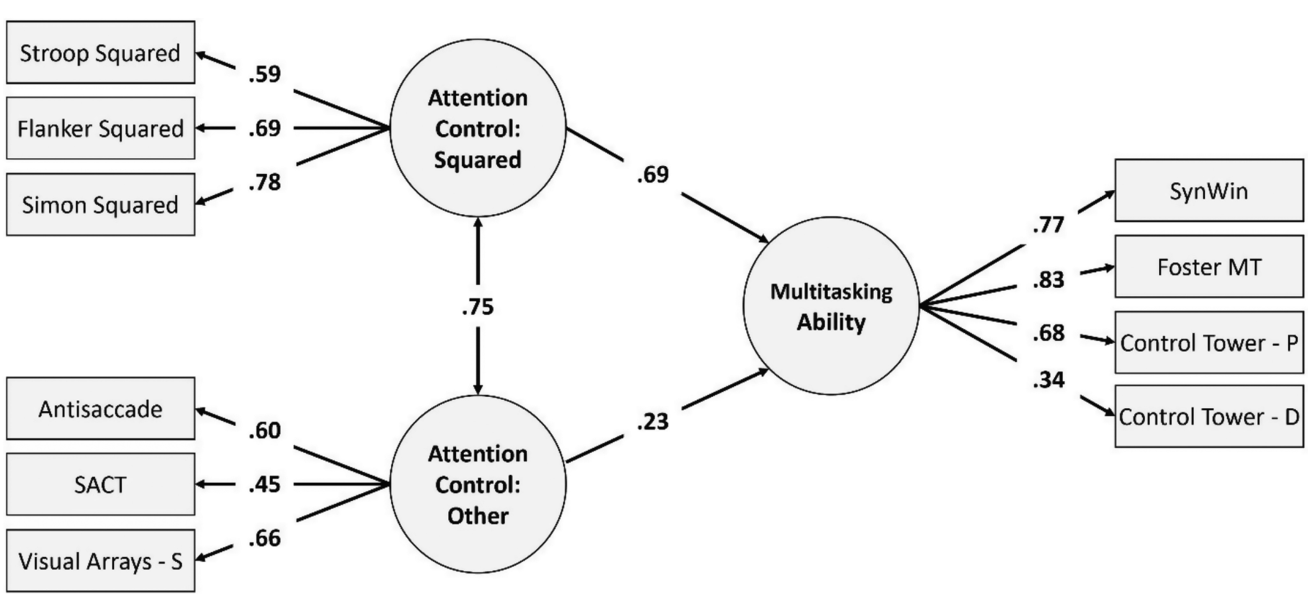
Accounting for the Positive Manifold

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

Figure 24

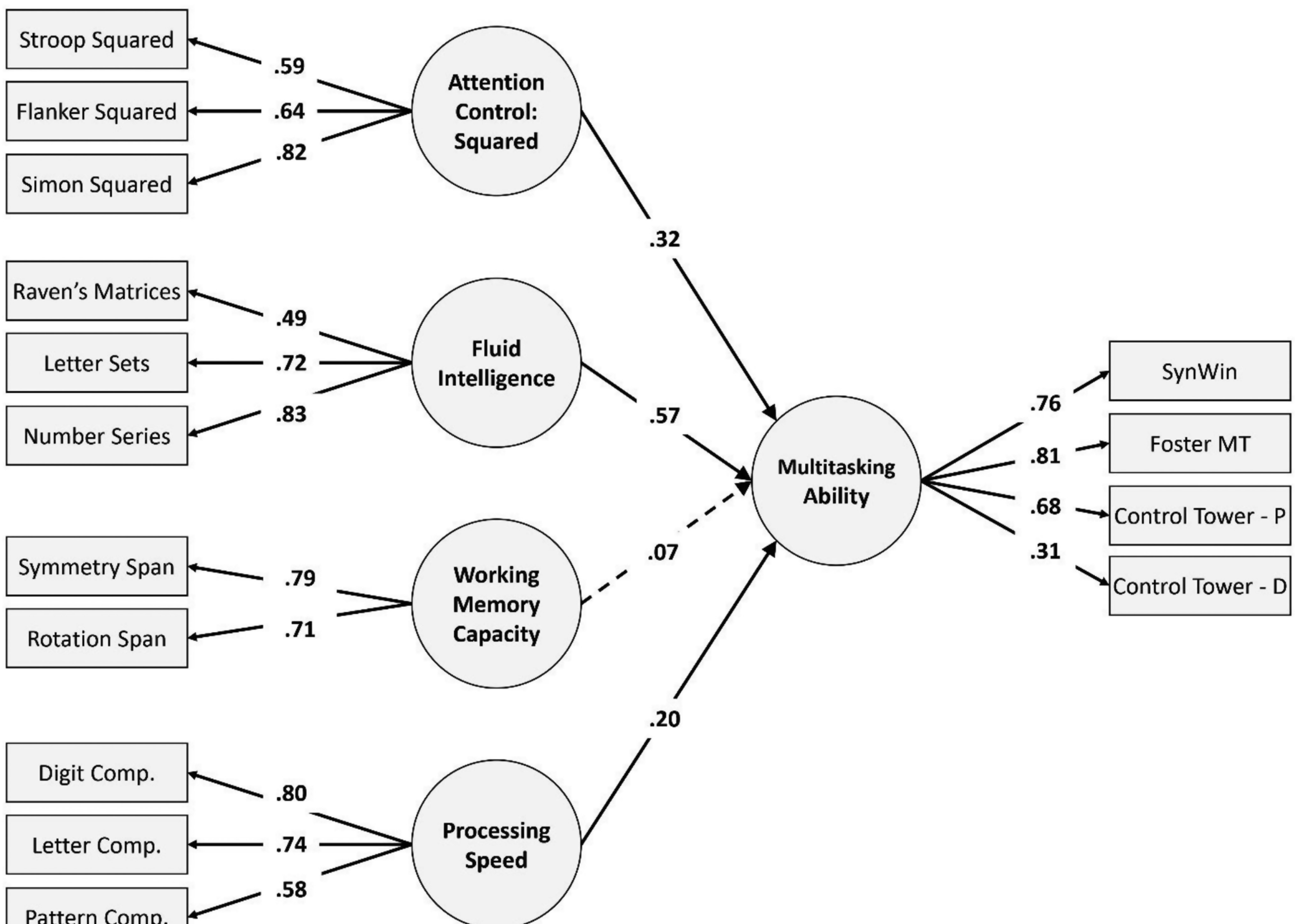
Structural Equation Model With the Squared Attention Control Factor and the Other Attention Control Factor Predicting Multitasking Ability



Note. $\chi^2(32) = 64.41$, $p < .001$; CFI = 0.950, TLI = 0.930, RMSEA = 0.069, 90% CI [0.047, 0.092], SRMR = 0.049.

Figure 25

Structural Equation Model With Squared Attention Control, Fluid Intelligence, Working Memory Capacity, and Processing Speed Predicting Multitasking Ability



Note. 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, $r = .44$). $\chi^2(80) = 163.20$, $p < .001$; CFI = 0.938, TLI = 0.919, RMSEA = 0.068, 90% CI [0.053, 0.082], SRMR = 0.057. AC = attention control; Gf = fluid intelligence; WMC = working memory capacity; PS = processing speed.

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.

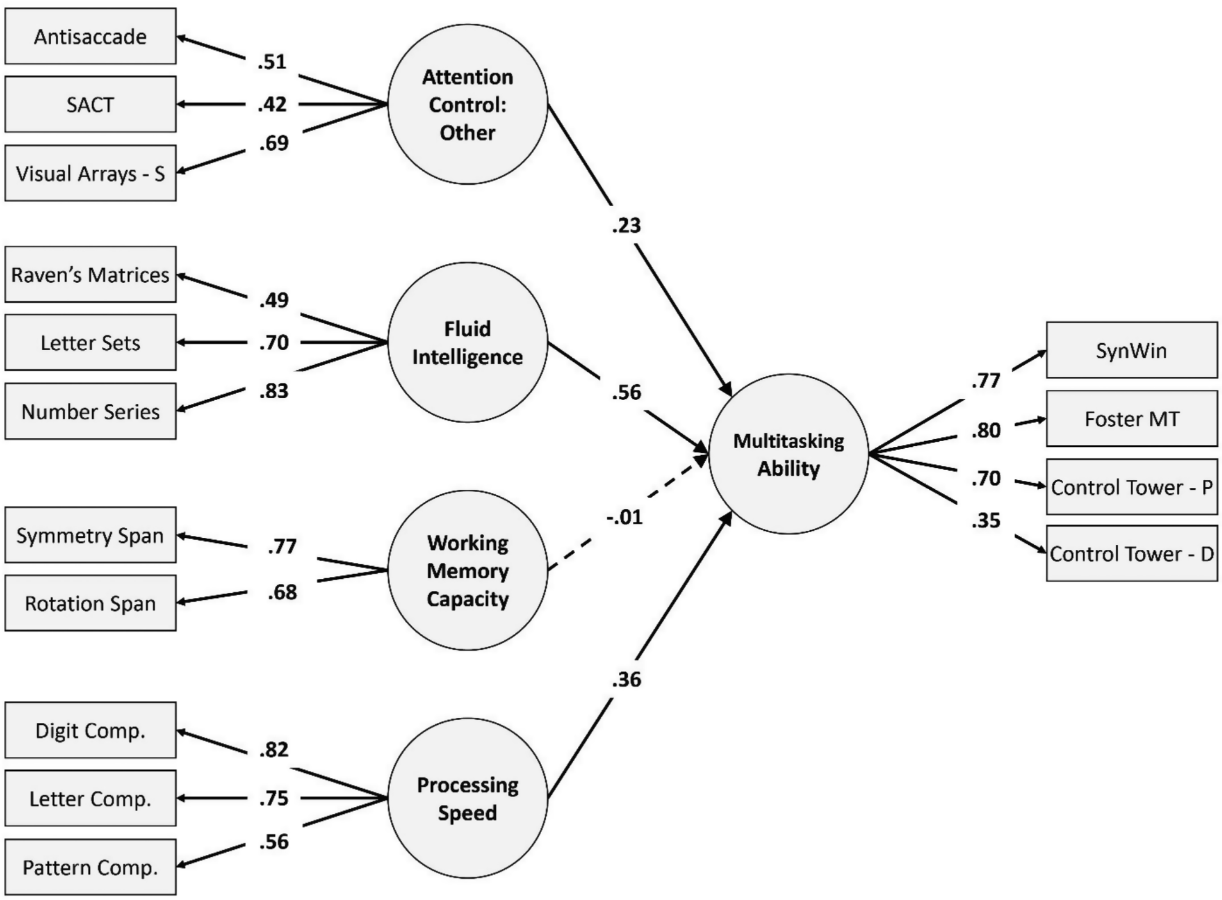
The broader purpose of these analyses was to determine the extent to which attention control explains the *positive manifold*—the positive correlations observed among broad cognitive abilities. We have argued that attention control is a domain-general ability that is required by a wide range of cognitive 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

Figure 26

Structural Equation Model With Non-Squared Attention Control, Fluid Intelligence, Working Memory Capacity, and Processing Speed Predicting Multitasking Ability



Note. 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, $r = .44$). $\chi^2(80) = 144.95$, $p < .001$; CFI = 0.944, TLI = 0.927, RMSEA = 0.060, 90% CI [0.044, 0.075], SRMR = 0.055. AC = attention control; Gf = fluid intelligence; WMC = working memory capacity; PS = processing speed.

around 55% of the variance. 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 min, amounting to 6 min of total testing time for the average participant. For comparison, the best three tasks from our lab's "toolbox" paper each required 12.5 min, on average, amounting to 37.5 min 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 10 min, 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 & Engle, 2020; Burgoyne et al., 2022; Engle, 2018, 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 maintaining 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.

Constraints on Generality

Across two studies, our sample included more than 600 individuals ages 18–35 recruited online across the United States (Study 1) and in the greater Atlanta, Georgia community (Study 2). Our conclusions are likely to be most applicable to samples of a similar age range, educational background, and level of English proficiency. Further validation is warranted for samples of children, adolescents, and older adults, as well as for nonnative English speakers and individuals with neurological disorders.

Context of Research

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., the Stroop task) 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.[1] 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. Taken together, this work shows that individual differences attention control can be reliably measured and contribute substantially to complex cognitive task performance.

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. https://doi.org/10.1037/0033-2909.131.1.30

Ackerman, P. L., & Hambrick, D. Z. (2020). A primer on assessing intelligence in laboratory studies. *Intelligence*, *80*, Article 101440. https://doi.org/10.1016/j.intell.2020.101440

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), 446–458. https://doi.org/10.1037/edu0000296

Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT battery. *Military Psychology*, *9*(1), 5–37. https://doi.org/10.1207/s15327876mp0901_1

[1] The three Squared tasks are freely available online: https://osf.io/7q598/.

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*, Article 386. https://doi.org/10.3389/fnins.2016.00386

Baddeley, A. D. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 5–28. https://doi.org/10.1080/713755608

Baumeister, R. F., Schmeichel, B. J., & Vohs, K. D. (2013). Self-regulation and the executive function: The self as controlling agent. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (pp. 516–539). Guilford Publications.

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. https://doi.org/10.1016/j.lindif.2011.01.007

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Burgoyne, A. P., & Engle, R. W. (2020). Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science*, *29*(6), 624–630. https://doi.org/10.1177/0963721420969371

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. https://doi.org/10.3758/s13423-019-01606-9

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. https://doi.org/10.1007/s00426-020-01348-7

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*, Article 101629. https://doi.org/10.1016/j.intell.2022.101629

Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., Hambrick, D. Z., & Engle, R. W. (2023). Understanding the relationship between rationality and intelligence: A latent-variable approach. *Thinking & Reasoning*, *29*(1), 1–42. https://doi.org/10.1080/13546783.2021.2008003

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). *Open data for nature and measurement of attention control*. Open Science Framework. https://doi.org/10.17605/OSF.IO/ZKQBS

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., Therriault, 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), 163–183. 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*(1), 42–100. https://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*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Donders, F. C. (1868). Over de snelheid van psychische processen. In *Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool (1868–1869)* (Series 2, pp. 92–120). https://doi.org/10.1016/0001-6918(69)90065-1

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. https://doi.org/10.1177/1745691615596990

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., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review*, *29*(4), 1143–1197. https://doi.org/10.3758/s13423-021-02052-2

Draheim, C., Tsukahara, J. S., & Engle, R. W. (2022, October 22). *Replication and extension of the toolbox approach to measuring attention control*. https://doi.org/10.31234/osf.io/gbnzh

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, *150*(2), 242–275. https://doi.org/10.1037/xge0000783

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests: 1976*. Educational Testing Service.

Elsmore, T. F. (1994). SYNWORK1: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instruments, & Computers*, *26*(4), 421–426. https://doi.org/10.3758/BF03204659

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*(2), 190–193. https://doi.org/10.1177/1745691617720478

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., & 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. In P. Jolicoeur, C. Lefebvre, & J. Martinez-Trujillo (Eds.), *Mechanisms of sensory working memory: Attention and perfomance XXV* (pp. 105–119). Academic Press.

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. https://doi.org/10.1098/rspl.1888.0082

Haaf, J. M., & Rouder, J. N. (2017). Developing constraint in Bayesian mixed models. *Psychological Methods*, *22*(4), 779–798. https://doi.org/10.1037/met0000156

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. https://doi.org/10.1080/14768320701212099

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279–1296. https://doi.org/10.1016/0042-6989(78)90218-3

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

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. https://doi.org/10.1080/10705519909540118

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), 851–860. https://doi.org/10.1037/a0023437

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*(3), 227–229. https://doi.org/10.3758/BF03214307

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

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. https://doi.org/10.1037/0096-3445.132.1.47

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), 127–143. https://doi.org/10.1037/neu0000598

Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

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), 2207–2249. https://doi.org/10.1037/xge0000774

Lezak, M. D. (1982). The problem of assessing executive functions. *International Journal of Psychology*, *17*(1-4), 281–297. https://doi.org/10.1080/00207598208247445

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. https://doi.org/10.1038/36846

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. D., Shipstead, Z., Harrison, T. L., Redick, T. S., Bunting, M., & Engle, R. W. (2020). The role of maintenance and disengagement in predicting reading comprehension and vocabulary learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(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*, *150*(12), 2525–2551. https://doi.org/10.1037/xge0001048

Martin, J. D., Mashburn, C. A., & Engle, R. W. (2020). 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. https://doi.org/10.1037/h0101851

McCabe, D. P., Roediger H. L., III, 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), 222–243. https://doi.org/10.1037/a0017619

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. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

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

Psychology Software Tools, Inc. [*E-Prime Go*]. (2020). 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), 127–140. https://doi.org/10.1037/0096-1523.26.1.127

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

Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., Kane, M. J., Hambrick, D. Z., & 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), 1473–1492. https://doi.org/10.1037/xge0000219

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. https://doi.org/10.1080/20445911.2012.704359

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), 501–526. https://doi.org/10.1037/xlm0000450

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763–776. https://doi.org/10.1037/0012-1649.27.5.763

Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin & Review*, *15*(2), 364–371. https://doi.org/10.3758/PBR.15.2.364

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/.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. https://doi.org/10.1016/j.jrp.2013.05.009

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. https://doi.org/10.1177/1745691616650647

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/.org/10.1037/h0020586

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

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*, *1*, ix+121.

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. https://doi.org/10.3758/s13414-020-02044-9

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

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Wiley, J., & Jarosz, A. F. (2012). How working memory capacity affects problem solving. In B. H. Ross (Ed.), *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. https://doi.org/10.1016/j.jecp.2010.04.007

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. https://doi.org/10.1111/mono.12032

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.

Received August 1, 2022
Revision received February 6, 2023
Accepted February 15, 2023 ■