Intelligence 104 (2024) 101835

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Knowledge, attention, and psychomotor ability: A latent variable approach to understanding individual differences in simulated work performance

Cody A. Mashburn[a,*], Alexander P. Burgoyne[b], Jason S. Tsukahara[b], Richard Pak[c], Joseph T. Coyne[d], Ciara Sibley[d], Cyrus Foroughi[d], Randall W. Engle[b]

[a] Kennesaw State University, Kennesaw, GA, USA
[b] Georgia Institute of Technology, Atlanta, GA, USA
[c] Clemson University, Clemson, SC, USA
[d] Information Technology Division, Naval Research Laboratory, Washington, DC, USA

ARTICLE INFO

Keywords:
Attention control
Multitasking
Crystallized intelligence
Psychomotor ability
personnel selection.

ABSTRACT

We compare the validity of personnel selection measures and novel tests of attention control for explaining individual differences in synthetic work performance, which required participants to monitor and complete multiple ongoing tasks. In Study 1, an online sample of young adults (N = 474, aged 18–35) based in the United States completed three-minute tests of attention control and two tests that primarily measure acquired knowledge, the Wonderlic and the Armed Forces Qualification Test (AFQT). Structural equation modeling revealed that acquired knowledge tests did not predict simulated work performance beyond attention control, whereas attention control did predict simulated work performance controlling for other measures. In Study 2, an in-lab sample of young adults from Georgia Tech and the greater Atlanta community (N = 321, aged 18–35) completed tests of attention control, processing speed, working memory capacity, and versions of two U.S. Military selection tests, one assessing acquired knowledge (the AFQT) and one assessing psychomotor ability (the Performance-Based Measures assessment from the Aviation Selection Test Battery). Structural equation modeling revealed that attention control fully mediated the relationship between the Performance Based Measures and simulated work performance, but the AFQT and processing speed retained unique prediction. We also explore possible gender differences. Collectively, these results suggest that tests of attention control may be a useful supplement to existing personnel selection measures when complex cognitive tasks are the criterion variable of interest.

1. Introduction

What underlies individual differences in work performance? Many studies have explored the role of acquired knowledge, as measured by selection tests such as the Wonderlic and the Armed Forces Qualification Test (Furnham, 2012; Hunter, 2017; Rakhmanov & Dane, 2021; Schmidt, Hunter, & Outerbridge, 1986; Schmidt, Hunter, Outerbridge, & Goff, 1988). Although these tests are often described as measures of "general cognitive ability" (see, e.g., Berry, Gruys, & Sackett, 2006; LePine, Colquitt, & Erez, 2000), the Wonderlic and Armed Forces Qualification Test are perhaps better described as domain-specific tests of verbal and numerical knowledge (e.g., Roberts et al., 2000). Nevertheless, domain-specific knowledge is important for many occupations. For instance, the Armed Forces Qualification Test is used across the United States Military because it is highly predictive of military training success and job performance (Schmidt & Hunter, 2004).

However valid knowledge tests may be, there are other sources of individual variation in cognition that might also predict vocational outcomes (McGrew, 2009). Recognizing that the exclusive use of knowledge tests likely overlooks useful information for predicting complex behaviors (e.g., flying a plane), researchers have also used psychomotor ability tests to predict performance (Fatolitis, Jentsch, Hancock, Kennedy, & Bowers, 2010; Fleishman, Teichner, & Stephenson, 1970; Fleishman, 1972; Fleishman & Rich, 1963; Gibb & Dolgin, 1988; Melton, 1947; Nye et al., 2020).

The U.S. military has explored several psychomotor tests over the decades (Fatolitis et al., 2010; Nye et al., 2020). For example, the Aviation Selection Test Battery, which is used by the U.S. Navy, Marine

* Corresponding author.
E-mail address: cmashb10@kennesaw.edu (C.A. Mashburn).

https://doi.org/10.1016/j.intell.2024.101835
Received 4 October 2023; Received in revised form 17 April 2024; Accepted 28 April 2024
Available online 8 May 2024

Corps, and Coast Guard to identify candidates for officer aviation training programs, added the Performance Based Measures when the test switched to computer-based administration in 2013 (Fatolitis et al., 2010; Walker, Olde, & Olson, 2007). The Performance Based Measures includes unimanual and bimanual tracking tasks, tests of selective attention, divided attention, and spatial/mental rotation. In a sample of Navy student pilots, Nye et al. (2020) found that the Performance Based Measures accounted for significant variance in flight performance above and beyond a non-computerized version of the Aviation Selection Test Battery, a composite score that, at the time of Nye et al.'s data collection, was based on acquired knowledge tests.

Although domain-specific acquired knowledge and psychomotor skill have been useful for predicting performance in various applied contexts, other, perhaps more fundamental, abilities should be considered as well. We suggest that *attention control,* the domain-general ability to maintain focus on task-relevant information and resist distraction and interference (Burgoyne & Engle, 2020), warrants particular consideration. Distractions are ubiquitous in the modern world. Attention control allows individuals to ignore or suppress the influence of these distractors to stay on task. Attention control is also implicated in most tasks that require controlled cognition, including performance on tests of working memory capacity, or the amount of task-relevant information that can be held active in memory (Engle, 2018). Specifically, attention control can be seen as one of the primary "active ingredients" that drives working memory capacity's predictive power (see, e.g., Engle, Tuholski, Laughlin, & Conway, 1999; Burgoyne, Mashburn, Tsukahara, & Engle, 2022). Accordingly, individual differences in attention control predict numerous real-world outcomes, including academic achievement (Ahmed, Tang, Waters, & Davis-Kean, 2019; Best, Miller, & Naglieri, 2011), job performance (Burgoyne et al., n.d., under review; Bosco, Allen, & Singh, 2015), and emotion regulation (Baumeister, Schmeichel, & Vohs, 2007; Schmeichel & Demaree, 2010; Zelazo & Cunningham, 2007); see Mashburn, Burgoyne, and Engle (2023) and Draheim, Pak, Draheim, and Engle (2022) for reviews. By measuring attention control alongside measures of acquired knowledge and psychomotor skill, the relative contribution of these cognitive abilities to job-relevant performance can be estimated.

1.1. Attention control and acquired knowledge

Ascertaining the relative contributions of these abilities is complicated by the fact that measures of acquired knowledge may share variance with measures of more "fluid" abilities, such as reasoning, problem-solving, attention, and memory. Specifically, it has been argued that the "investment" of fluid abilities results in the acquisition of crystallized knowledge (Cattell, 1963; Schweizer & Koch, 2002). Indeed, research indicates that knowledge acquisition relies on fluid cognitive abilities that contribute to learning, sometimes over years of instruction and study (Cattell, 1963; Kvist & Gustafsson, 2008). Successfully applying these fluid abilities should lead to greater knowledge acquisition, so knowledge tests may tap fluid abilities by proxy (Cattell, 1987; Ohi et al., 2022).

In previous research, we found that the Armed Services Vocational Aptitude Battery, which is largely a test of acquired knowledge (Roberts et al., 2000), strongly predicted performance in simulated work multitasking paradigms (Martin, Mashburn, & Engle, 2020). Accounting for attention control and fluid intelligence, however, nullified this relationship. That is, when attention control, fluid intelligence, and the Armed Services Vocational Aptitude Battery were specified as predictors of multitasking performance, the unique contribution of the Armed Services Vocational Aptitude Battery was small and nonsignificant. By contrast, the unique contributions of attention control and fluid intelligence were large and significant, suggesting that the Armed Services Vocational Aptitude Battery predicted multitasking only because it was related to attention control and fluid intelligence.

Similar results have been obtained for the Wonderlic. Hicks, Harrison, and Engle (2015) administered the Wonderlic alongside tests of working memory capacity (which demands both controlled attention and short-term storage of information; see Engle et al., 1999) and fluid intelligence. Although the Wonderlic correlated strongly with fluid intelligence, controlling for working memory capacity fully accounted for the relationship between Wonderlic performance and fluid reasoning ability. Bosco et al. (2015) found complimentary results outside of the laboratory; they administered the Wonderlic and tests of attentional abilities to a sample of private-sector financial workers. A composite of their attention tests explained significant variance in supervisor ratings and performance on a simulated management task above and beyond the Wonderlic. Not only is attention control fundamental to complex cognition, explaining variance that might otherwise be attributed to acquired knowledge, it predicts real-world behavior above and beyond general knowledge tests.

1.2. Attention control and psychomotor ability

Similarly, performance on various psychomotor tasks may also depend in part on attention control. For example, one cannot complete a psychomotor task such as unimanual tracking without attending to the target, attending to the tracking crosshair, and continuously monitoring performance to reduce the distance between the target and the crosshair. All of these aspects of task performance point to an important role for controlled attention, and some researchers have even employed such continuous tracking tasks as markers of attentional lapses (Unsworth, Robison, & Miller, 2021). It is somewhat surprising, then, that not much research has investigated whether individual differences in attention control underpin individual differences in psychomotor performance. In part, this may be due to the recent development of more reliable tests of attention control, on which the present study capitalizes (e.g., Burgoyne, Tsukahara, Mashburn, Pak, & Engle, 2023; Draheim, Tsukahara, Martin, Mashburn, & Engle, 2021).

1.3. Research objectives

Our aim is to compare the predictive validity of existing selection measures with attention control. Specifically, we aim to compare the validity of knowledge measures, psychomotor ability measures, and a set of novel attention control measures for predicting simulated work performance, which requires participants to keep track of and complete multiple ongoing tasks (see Barron & Rose, 2017). Simulated work or "synthetic work" tasks provide a means of studying work performance in the laboratory. They share many demands with many real-world work scenarios, but purposefully do not strongly resemble any particular job (Hambrick, Burgoyne, Altmann, & Matteson, 2023). The intention behind these tasks is to capture the cognitive demands common to a wide range of work tasks without tapping occupational-specific knowledge that might afford differential advantages to those with more experience in a domain. To the extent that this is true, using synthetic work tasks as criterion measures allows researchers to generalize laboratory-based findings to draw conclusions more broadly about work performance (at least for occupations which require managing multiple ongoing events) than any one "job simulator" task might afford.

We expected attention control to explain unique variance in simulated work performance controlling for the other measures. Moreover, we expected attention control to help explain the association between acquired knowledge/psychomotor ability and simulated work performance. Study 1 evaluated these claims with a particular focus on acquired knowledge tests in an online sample. Study 2 was conducted in-lab to assuage concerns about the online administration of knowledge tests and also included a psychomotor ability assessment, the Performance Based Measures from the U.S. Navy's Aviation Selection Test Battery. Study 2 also evaluated other explanatory alternatives to attention control, including working memory capacity and processing speed. Ultimately, we aim to encourage applied researchers and practitioners to consider tests of fluid abilities, and attention control in particular, to increase the validity, and possibly equity, of their personnel selection practices when complex cognition is required for performing well in a position.

2. Study 1

In Study 1, we compared the prediction of simulated work multitasking performance by novel tests of attention control (i.e., the three-minute "Squared" tasks, Burgoyne, Tsukahara, Mashburn, Pak, & Engle, 2023) and two existing measures of acquired knowledge, the Armed Forces Qualification Test and the Wonderlic.

2.1. Method

2.1.1. Research ethics and consent

This study was approved by the Georgia Institute of Technology's Institutional Review Board under Protocol H21393. All participants provided written consent via a digital consent form prior to participating in this study.

2.1.2. Participants

Our initial sample consisted of 516 respondents recruited through Prolific. Our recruitment filters required participants to be ages 18–35, based in the United States, native English speakers, have normal or corrected-to-normal vision, and have no history of seizures due to the nature of some of the test stimuli. Additional participants were screened out due to unacceptably low performance or outlying data points; this is detailed further in our Data Preparation section. Participants were credited $30 for completing the study.

Demographic information for the 474 participants who provided data on the Wonderlic is provided in Table 1. The participants ranged in age from 18 to 35; the mean age was 27.5 years. Approximately 80% of the sample had attended or were attending college. Approximately one-third of the sample identified as Black or African American, 31% identified as White, 22% identified as Hispanic or Latino, and 14% selected more than one racial/ethnic category. Additionally, while most participants reported completing at least some college course work (79.5%), a sizeable minority report no experience with higher education (20.5%). As such, our sample presents a broad array of demographic and sociocultural experience, which increases external validity to the larger United States population.

2.1.3. Procedure

Participants were directed from Prolific to a Qualtrics survey where they were assigned a unique participant number. On Qualtrics, they completed the Armed Forces Qualification Test and the Wonderlic test, described below. Upon completing these multiple-choice tests, they followed an external link to a study dashboard providing access to the remaining tasks in the study: Stroop Squared, Flanker Squared, Simon Squared, and SimWork. Once participants completed the three Squared tests of attention control, they completed SimWork. Finally, they followed an external link back to Prolific where they attested to completing the study and were awarded credit. The median testing time was approximately 1.5 h.

Table 1
Demographic information for Study 1.

Demographic	Statistic
Age	Mean: 27.5 *SD*: 4.8 Range: 18–35
Gender	Female: 38.6% Male: 57.8% Transgender: 0.8% Other/Self-Identify: 2.7%
At least some college?	Yes: 79.5% No: 20.5%
Race/Ethnicity	White: 31.0% Black or African American: 32.9% Hispanic or Latino: 21.7% Mixed/Selected more than one category: 14.4%

$N = 474$.

2.1.3.1. Armed forces qualification test. Participants completed four multiple-choice subtests designed to measure arithmetic reasoning, word knowledge, paragraph comprehension, and mathematics knowledge. Because the Armed Forces Qualification Test is proprietary, items were selected from a practice test book (Powers, 2011) on the basis that they appeared to represent a range of difficulty and content areas. For all subtests, items that were not completed in the allotted time were scored as incorrect. The Armed Forces Qualification Test is normally scored as a weighted sum of its four subtests (described below). The formula is:

$$AFQT\ Score = 2^{*}(Word\ Knowledge + Pragraph\ Comprehension) + Arithmetics\ Reasoning + Mathematics\ Knowledge \quad (1)$$

While many of our analyses involve the subtests directly, we also report descriptive statistics and bivariate correlations for this composite score.

2.1.3.1.1. Arithmetic reasoning (AR). Participants were asked to complete 24 items in 28.8 min. An example item is "Your piggy bank contains $19.75 in dimes and quarters. There are 100 coins in all. How many dimes are there?" (response options: 25, 30, 35, 40).

2.1.3.1.2. Word knowledge (WK). Participants were asked to complete 28 items in 8.8 min. An example item is "Abeyance most nearly means ___" (response options: trustworthiness, passion, suspension, business).

2.1.3.1.3. Paragraph comprehension (PC). Participants were presented 12 short passages of text and asked one question per passage. They were given 10.4 min for this subtest. Passage topics ranged from science and mechanical engineering to social studies and politics. Some questions asked about factual statements made in the passage of text, whereas others asked participants to make a straightforward inference given the content of the passage.

2.1.3.1.4. Mathematics knowledge (MK). Participants were asked to complete 20 items in 19.2 min. An example item is "The cube of 5 is ___" (response options: 125, 25, 15, 50).

2.1.3.2. Wonderlic. Participants were given 12 min to complete a Wonderlic practice test comprising 50 multiple-choice questions assessing general knowledge, algebra, geometry, vocabulary, clerical, and spatial reasoning abilities. Because the Wonderlic is proprietary, questions were sampled from an online training resource (wonderlictestpractice.com). An example item is "A girl is 21 years old and her brother is a third her age. When the girl is 36, what will be the age of her brother?" (response options: 12, 22, 27, 17). Another example item is "Are the following two words [Legend, Key] similar, contradictory, or not related?" Due to errors in the source material, two items administered to participants had no correct response options; these two items were removed from analysis. The outcome measure was the number of correct responses, with a maximum possible score of 48.

2.1.3.3. Attention control. Participants completed the three-minute "Squared" tests as measures of attention control (Burgoyne, Tsukahara, Mashburn, Pak, & Engle, 2023). These tests build on classic experimental paradigms (e.g., the Stroop, flanker, and Simon tasks) but add an additional twist. As described below, conflict can occur at the level of the stimulus as well as at the level of the response options. Subjects must pay attention to one attribute (or characteristic) of the target stimulus and a different attribute or characteristic when considering the response options. Thus, the tasks also involve a switching element whereby the stimulus dimension that was irrelevant when considering the target stimulus becomes relevant when making a response.

2.1.3.3.1. Stroop squared (Burgoyne, Tsukahara, Mashburn, Pak, & Engle, 2023). In Stroop Squared (Fig. 1), 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. After reading the instructions, participants were given 30s of practice with auditory (a short auditory chime for correct items or a buzzer sound for errors) and visual feedback ("+1 point" for correct answers, "-1 point" for errors) on each trial. Participants also saw their current point total and the amount of time remaining displayed on screen. After 30 s of practice, participants reviewed the instructions again before doing the task continuously for 90 s. The measure of performance was the number of correct responses minus the number of incorrect responses.

2.1.3.3.2. Flanker Squared (Burgoyne, Tsukahara, Mashburn, Pak, & Engle, 2023). In Flanker Squared (Fig. 2), 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., > > < > >). After reading the instructions, participants were given 30s of practice with auditory (a short auditory chime for correct items or a buzzer sound for errors) and visual feedback ("+1 point" for correct answers, "-1 point" for errors) on each trial. Participants also saw their current point total and the amount of time remaining displayed on screen. After 30 s of practice, participants reviewed the instructions again before doing the task continuously for 90 s. The measure of performance was the number of correct responses minus the number of incorrect responses.

2.1.3.3.3. Simon Squared (Burgoyne, Tsukahara, Mashburn, Pak, & Engle, 2023). In Simon Squared (Fig. 3), 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." After reading the instructions, participants were given 30s of practice with auditory (a short auditory chime for correct items or a buzzer sound for errors) and visual feedback ("+1 point" for correct answers, "-1 point" for errors) on each trial. Participants also saw their current point total and the amount of time remaining displayed on screen. After the 30 s practice, participants reviewed the instructions again before doing the task continuously for 90 s. The measure of performance was the number of correct responses minus the number of incorrect responses.

2.1.3.4. SimWork. In this simulated work multitasking paradigm, participants were challenged to complete four subtasks concurrently (Fig. 4). This task was modeled after SynWin (Elsmore, 1994), with the exception that the auditory task used in SynWin was replaced with a visual monitoring task to facilitate online testing. First, participants were given one minute of practice on each of the four subtasks, one at a time. Next, they completed three five-minute blocks of testing during which they attempted to complete all four subtasks concurrently to maximize their score.

The first subtask was a Sternberg memory probe in which participants were briefly shown 7 letters to memorize. Every 20 s a letter would appear for 5 s and the participant's task was to determine whether it was one of the 6 letters in the memory set or not by clicking one of two buttons. Participants could click the memory set box at any time to display the 7 letters again, however, this would subtract points from their score. Participants were awarded 10 points for each correct response and lost 10 points for each incorrect response, failure to respond, or re-display of the memory set.

The second subtask was a fuel gauge monitoring task. Participants were shown a fuel gauge that steadily decreased over time. They could reset the fuel gauge by clicking on it. Participants were awarded maximum points (10 points) for resetting the fuel gauge when it was as low as possible, and fewer points for resetting the fuel gauge when it had more fuel remaining. They lost 10 points for every second that the fuel gauge remained empty.

The third subtask was a mathematics task. Participants were challenged to add two three-digit numbers together, using the mouse to indicate their response. Participants were awarded 10 points for each item that they solve correctly, and lost 10 points for each incorrect response.

The fourth subtask was a shape monitoring task. Participants were shown a grid of squares, and every 5 s one of the squares would rotate 45 degrees to a "diamond" orientation. Participants were instructed to click on the diamond to "reset" it to a square shape. Participants earned 10 points for clicking on the diamond within 3 s. If no response was made, participants lost 5 points per second until they responded.

2.1.4. Data preparation

To maximize the validity of our analyses, 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. For the Armed Forces Qualification Test subtests and the Wonderlic, scores of zero were set to missing. For the SimWork multitask and the attention control tasks, negative scores were set to missing. We also screened for outliers, which we define as any value with a standardized score (i.e., z-score) ± 3.5 standard deviations from the sample mean. After setting these outlying scores to missing, we recomputed standard deviations and re-checked for outliers. We repeated this process until no further outliers were identified. A summary of removed cases for Study 1 based on this data preparation process can be found in Table 2. In addition, some missing data occurred due to technical problems participants encountered during online study administration. Finally, we note that because the Armed Forces Qualification Test is a composite variable based on four subtests, a missing score on one subtest would render the composite uninterpretable. Therefore, if a participant had an outlying or missing score on any one subtest, we set all their other Armed Forces Qualification Test subtest scores to missing, resulting in a final sample of 451 cases for all Armed Forces Qualifying Test variables.

2.1.5. Modeling approach and fit statistics

We used maximum likelihood estimation with robust standard errors in JASP 0.17.1 for all confirmatory factor analyses and structural equation models; missing data were handled using full information maximum likelihood estimation (JASP Team, 2023). Variables were standardized before entry into any 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. The comparative fit index (CFI) and Tucker-Lewis index (TLI) 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., > 0.90 or ideally, > 0.95). For the root mean square error of approximation (RMSEA) fit statistic, values <0.05 are considered great, but values <0.10 are acceptable. For the standardized root mean square residual (SRMR), which computes the standardized difference between the observed and predicted correlations, a value of <0.08 indicates good fit (Hu & Bentler, 1999).

2.2. Results

2.2.1. Descriptive statistics and reliability

Descriptive statistics are provided in Table 3. The three Squared tests of attention control had split-half reliability estimates of 0.79, 0.84, and 0.85, and the attention control composite score had a Cronbach's alpha

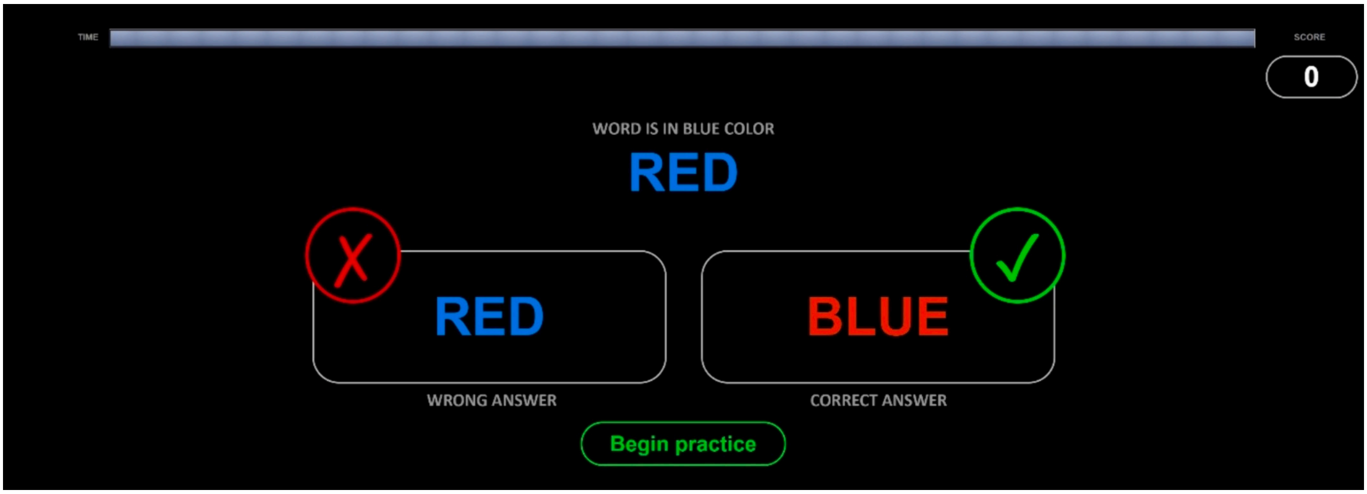

Fig. 1. The stroop squared instruction screen.
The participant's task is to select the response option with the word meaning that matches the display color of the target stimulus. In the above example, the target stimulus is the word "RED" appearing with a blue display color, so the participant must select the response option that says the word "BLUE" (i.e., the one on the right).

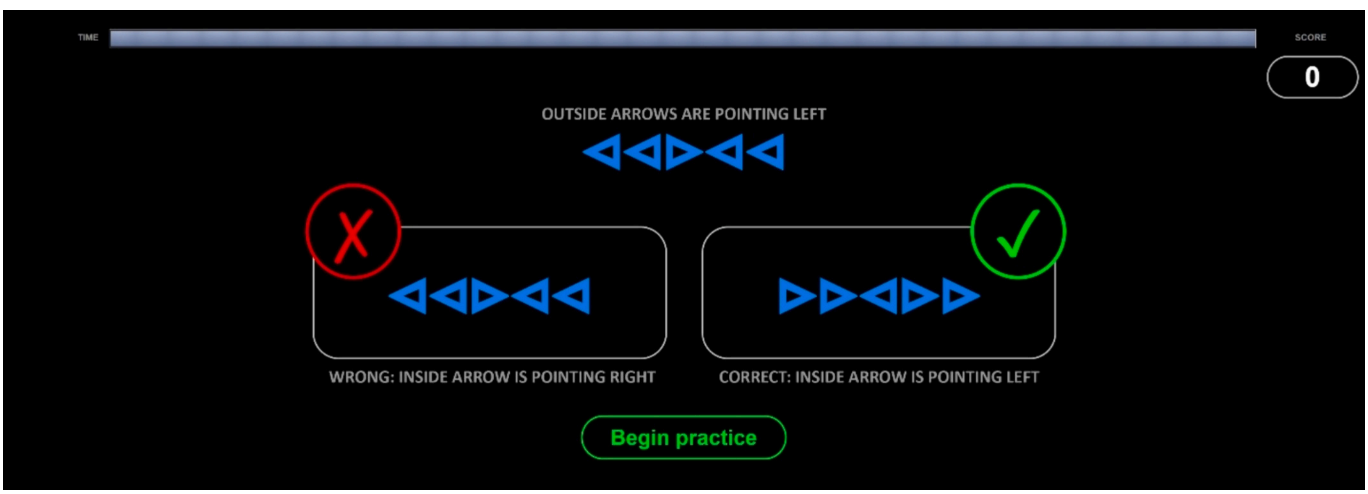

Fig. 2. The Flanker Squared instruction screen.
The participant's task is to select the response option with a central arrow that points in the same direction as the flanking arrows in the target stimulus. In the above example, the target stimulus has flanking arrows pointing left, so the participant must select the response option which has a central arrow pointing left.

of 0.76. The four subtests of the Armed Forces Qualification Test had internal consistencies ranging from 0.63 (paragraph comprehension) to 0.83 (mathematics knowledge) while the entire test had a Cronbach's alpha of 0.89.[1] The Wonderlic had a Cronbach's alpha of 0.90. Finally, Cronbach's alpha across the three blocks of SimWork multitasking was 0.84.

2.2.2. *Bivariate correlations*

All correlations between measures were computed with pairwise deletion and are presented in Table 4. We expected to see positive zero-order correlations among all variables. Indeed, all correlations were statistically significant at the $p < .01$ level. The three Squared tests of attention control had correlations with one another ranging from $r = 0.52$ to $r = 0.56$ (avg. $r = 0.54$), indicating that they tapped similar sources of variance. The correlation between the attention control tasks and average SimWork performance ranged from $r = 0.38$ to $r = 0.52$ (avg. $r = 0.46$). For comparison, the Wonderlic correlated $r = 0.35$ with average SimWork performance. The four subtests of the Armed Forces Qualification Test had correlations with average SimWork performance that ranged from $r = 0.23$ to $r = 0.36$ (avg. $r = 0.29$), while the Armed Forces Qualification Test composite score correlated at $r = 0.38$. We now turn to our confirmatory factor analysis and structural equation modeling analyses to compare the prediction of simulated work performance by the attention control and knowledge assessments.

2.2.3. *Structural equation modeling*

We used structural equation modeling to examine the relationships between attention control, the Wonderlic, the Armed Forces Qualification Test, and SimWork multitasking at the latent level. Structural equation modeling is a statistical technique for isolating variance common to a set of observed measures and then using that common variance to draw inferences about the relationships among unobserved, latent variables.

First, we conducted a confirmatory factor analysis to evaluate correlations between latent factors derived from our measures. We defined

[1] Cronbach's alpha for the Armed Forces Qualification Test is computed from participants' accuracy data across the entire set of items with no adjustment for subtest weighting.

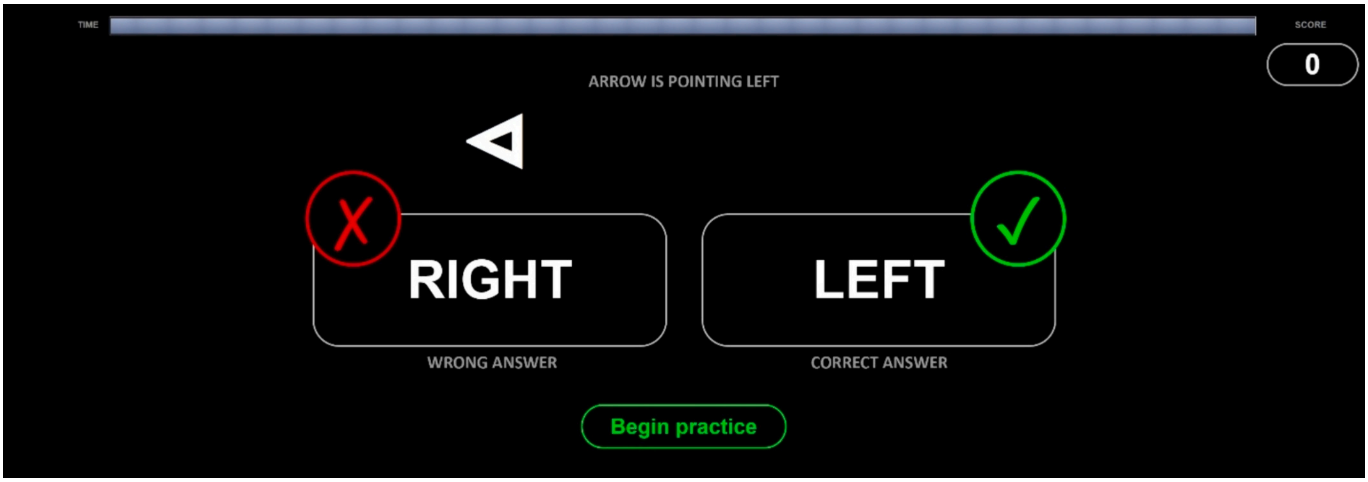

Fig. 3. The Simon Squared instruction screen.
The participant's task is to select the response option that states the direction that the arrow is pointing. In the above example, the arrow is pointing left, so the participant must select the response option that says "LEFT".

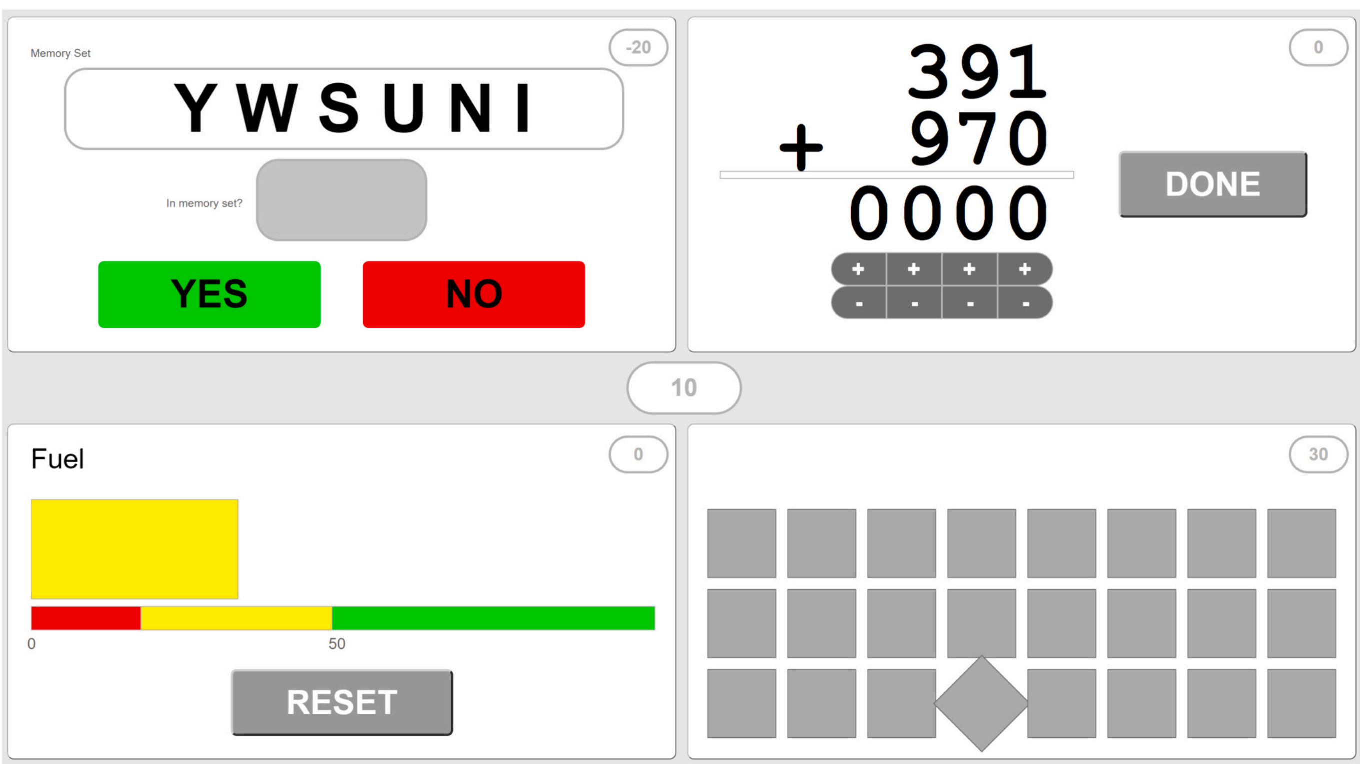

Fig. 4. The SimWork interface.
Participants must manage four subtasks concurrently to maximize their score. The four tasks are Sternberg memory probe, fuel gauge monitoring, math problems, and shape monitoring.

a latent factor representing attention control using the three Squared tests of attention control and defined a latent factor representing simulated work multitasking performance using the three blocks of SimWork. We modeled the Armed Forces Qualification Test and Wonderlic as latent variables to maintain comparability with attention control and simulated work performance. We formed an Armed Forces Qualification Test factor by loading each of the four subtests onto one latent variable. For the Wonderlic, we computed item-total correlations and divided the 48 items into three 16-item parcels with comparable average item-total correlations. The mean item-total correlation for items in the three parcels was 0.38, 0.39, and 0.38. We then summed item responses within each parcel and formed a latent Wonderlic factor from them. All latent factors were allowed to freely correlate; this model fit the data well, χ^2 (59) = 186.892, $p < .001$; CFI = 0.954, TLI = 0.939, RMSEA = 0.067, 90% CI [0.056, 0.078], SRMR = 0.044.

As shown in Table 5, correlations between latent factors representing attention control, the Wonderlic, and the Armed Forces Qualification Test were statistically significant and ranged from $r = 0.52$ to $r = 0.64$ (all $ps < 0.001$), indicating that the predictors share between 27.04% and 40.96% of their reliable variance (i.e., $.52^2$ and $.64^2$). All correlations with the SimWork factor were also large and significant ($ps <$ 0.001), ranging from $r = 0.74$ for attention control to $r = 0.50$ for the

Table 2
Summary of removed cases for Study 1.

Criterion	Task	Cases Removed	
Poor Performance			*Cutoff*
	Wonderlic	39	0 Points
	AFQT Arithmetic Reasoning	53	0 Points
	AFQT Word Knowledge	54	0 Points
	AFQT Paragraph Comprehension	58	0 Points
	AFQT Mathematics Knowledge	56	0 Points
	SimWork Block 1	37	< 0 Points
	SimWork Block 2	8	< 0 Points
	SimWork Block 3	6	< 0 Points
	Stroop Squared	40	< 0 Points
	Flanker Squared	42	< 0 Points
	Simon Squared	2	< 0 Points
Outlier (±3.5 SDs)			*Outlier Passes*
	AFQT Word Knowledge	6	2
	AFQT Paragraph Comprehension	1	1
	SimWork Block 1	1	1
	SimWork Block 3	4	3

AFQT = Armed Forces Qualification Test.

AFQT, to $r = 0.40$ for the Wonderlic.

Next, we specified a structural equation model to assess the unique contribution of the attention control, Armed Forces Qualification Test, and Wonderlic factors to the prediction of SimWork multitasking (see Fig. 5). Critically, the predictive path from attention control to multitasking ability was large and statistically significant ($\beta = 0.70, p < .001$), whereas the predictive paths from the Wonderlic ($\beta = 0.04, p = .46$) and the Armed Forces Qualification Test ($\beta = 0.03, p = .65$) were small and not statistically significant. Together, the predictors accounted for 54.7% of the variance in multitasking ability at the latent level. A second model revealed that, on its own, attention control accounted for 55.0% of the variance in multitasking ability at the latent level (see Appendix A for a depiction of this model). This indicates that attention control accounted for virtually all of the predictive validity of the model depicted in Fig. 5.

2.3. Summary of Study 1

Study 1 investigated the relative contribution of attention control and two knowledge-based tests (i.e., the Wonderlic and the Armed Forces Qualification Test) to performance in a simulated work multitasking paradigm. At the latent level, all predictors were significantly correlated with simulated work performance. However, structural equation modeling revealed that attention control accounted for a majority of the variance in simulated work performance on its own ($R^2 = 55.0\%$), and adding the Wonderlic and the Armed Forces Qualification Test did not increase the total R^2 (i.e., $R^2 = 54.7\%$). In the full model, only the contribution of attention control to simulated work multitasking performance was statistically significant ($\beta = 0.70, p < .001$).

3. Study 2

Study 1 revealed that attention control is a much stronger predictor of multitasking performance than the Armed Forces Qualification Test and the Wonderlic. This finding is consistent with some other studies in the literature. For example, Redick et al. (2016) found that attention control, working memory capacity, and fluid intelligence all strongly predicted multitasking ability. They also found that attention control helped explain the association between working memory capacity and multitasking performance. Similarly, Martin et al. (2020) found that the Armed Services Vocational Aptitude Battery (from which the Armed Forces Qualification Test is derived) only predicted multitasking when attention control and fluid intelligence were not controlled for.

However, some studies have reported a unique contribution of acquired knowledge to multitasking. For instance, Hambrick et al. (2011) found that the Armed Services Vocational Aptitude Battery predicted multitasking, although the effect was partly explained by memory updating performance. There are several pertinent differences between the design of Hambrick et al. (2011) and Study 1. One important difference is that Hambrick et al.'s (2011) participants were tested in a proctored environment, whereas our online sample was not.

This raises questions about the validity of the knowledge tests in Study 1. Many of the answers to the mathematics and vocabulary items in the Armed Forces Qualification Test and the Wonderlic can be found online. This is not a problem for the attention control tests, because participants cannot look up the answers, and attempting to do so would come at the cost of limited testing time (e.g., 90 s for each of the three Squared tests of attention control). Thus, cheating could have differentially affected the measurement of acquired knowledge relative to attention control. It is possible that in a proctored testing environment, acquired knowledge tests such as the Wonderlic and Armed Forces Qualification Test would be more powerful predictors of simulated work multitasking performance (c.f., Hambrick et al., 2011). Given concerns about the validity of online study data, we conducted a second study in our laboratory.

Study 1 was limited in other ways as well. Although attention control emerged as the strongest predictor of simulated work performance, it is one among many constructs that could serve as useful augmentations to

Table 3
Descriptive statistics for Study 1.

Measure	*N*	*M*	*SD*	Skew	Kurtosis	Reliability
Wonderlic Score	474	21.31	9.00	0.18	−0.47	0.90[α]
AFQT Arithmetic Reasoning	451	13.30	4.28	−0.02	−0.50	0.78[α]
AFQT Mathematics Knowledge	451	12.60	4.47	−0.29	−0.79	0.83[α]
AFQT Word Knowledge	451	20.52	3.31	−0.49	0.42	0.68[α]
AFQT Paragraph Comprehension	451	9.23	2.04	−0.87	0.56	0.63[α]
AFQT Composite Score	451	85.40	14.77	−0.46	0.02	0.89[α]
Stroop Squared	398	26.24	13.86	−0.36	−0.86	0.84[b]
Flanker Squared	388	22.24	12.92	−0.08	−0.72	0.85[b]
Simon Squared	396	47.22	13.84	−1.04	1.00	0.79[b]
SimWork Multitask Block 1	361	701.97	263.81	−0.46	−0.17	–
SimWork Multitask Block 2	361	830.48	221.06	−0.54	0.36	–
SimWork Multitask Block 3	361	847.91	212.79	−0.74	0.76	–
SimWork Multitasking Average	361	793.45	203.44	−0.45	0.15	0.84[α]

— Reliability could not be computed for the individual blocks of the SimWork multitask.
[α] Cronbach's alpha.
[b] Split-half reliability with Spearman-Brown correction.

Table 4
Correlations between major variables from Study 1.

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. SimWork Average	–											
2. SimWork Block 1	**0.87**	–										
3. Sim Work Block 2	**0.89**	**0.66**	–									
4. SimWork Block 3	**0.86**	**0.58**	**0.71**	–								
5. Stroop Squared	**0.41**	**0.38**	**0.39**	**0.28**	–							
6. Flanker Squared	**0.48**	**0.40**	**0.44**	**0.42**	**0.52**	–						
7. Simon Squared	**0.52**	**0.43**	**0.47**	**0.48**	**0.56**	**0.53**	–					
8. Wonderlic Score	**0.35**	**0.34**	**0.31**	**0.25**	**0.33**	**0.29**	**0.39**	–				
9. AFQT Composite	**0.39**	**0.36**	**0.37**	**0.29**	**0.37**	**0.44**	**0.36**	**0.43**	–			
10. AFQT AR	**0.36**	**0.33**	**0.32**	**0.28**	**0.39**	**0.44**	**0.41**	**0.40**	**0.85**	–		
11. AFQT MK	**0.30**	**0.28**	**0.31**	**0.20**	**0.33**	**0.37**	**0.25**	**0.36**	**0.86**	**0.67**	–	
12. AFQT WK	**0.23**	**0.19**	**0.22**	**0.19**	**0.14**	**0.20**	**0.16**	**0.29**	**0.67**	**0.35**	**0.40**	–
13. AFQT PC	**0.25**	**0.25**	**0.22**	**0.17**	**0.31**	**0.36**	**0.34**	**0.28**	**0.64**	**0.47**	**0.38**	**0.39**

Boldface, statistically significant at $p < .01$. Pairwise n ranges from 333 to 451. AFQT = Armed Forces Qualification Test, AR = Arithmetic Reasoning, MK = Mathematical Knowledge, WK = Word Knowledge, PC = Paragraph Comprehension.

Table 5
Correlations between latent factors from Study 1.

Factor	1.	2.	3.
1. Multitasking (SimWork)	–		
2. Attention Control	**0.74**	–	
3. Armed Forces Qualification Test	**0.50**	**0.64**	–
4. Wonderlic	**0.40**	**0.49**	**0.52**

Bold, statistically significant at $p < .001$. All factor loadings were large and statistically significant. For the multitasking factor, loadings ranged from 0.76 (Block 1) to 0.89 (Block 2). The factor loadings for the three attention control tests ranged from 0.72 (Stroop Squared) to 0.77 (Simon Squared). For the Wonderlic, factor loadings ranged from 0.87 (Item Parcel 2) to 0.90 (Item Parcel 1). For the Armed Forces Qualification Test, factor loadings ranged from 0.48 (Word Knowledge) to 0.85 (Arithmetic Reasoning). Model fit was acceptable, χ^2 (59) = 186.892, $p < .001$; CFI = 0.954, TLI = 0.939, RMSEA = 0.067, 90% CI [0.056, 0.078], SRMR = 0.044.

existing selection instruments. Two others are working memory capacity and processing speed. Working memory capacity and attention control are closely related constructs. Engle and colleagues (Engle, 2002; Engle, 2018; Engle et al., 1999) have argued that “attention control” corresponds to the executive control component common to many models of working memory (e.g., Baddeley, 2012; Cowan, 1988; Cowan, Morey, & Naveh-Benjamin, 2020). Further, working memory capacity processes have been implicated in many situations requiring controlled attention, particularly when information maintenance, updating, or retrieval are required (Redick, Calvo, Gay, & Engle, 2011; Meier, Smeekens, Silvia, Kwapil, & Kane, 2018; Unsworth & Engle, 2007). Such demands appear to be present in our simulated work multitask from Study 1, as participants must strategically select among multiple competing sub-goals. Given their theoretical relatedness, it would be beneficial to know whether measures of working memory capacity and attention control independently predict simulated work performance. Additionally, there are clear benefits to faster cognitive processing in our attention control tasks and multitasks, given their timed nature. As such, individual

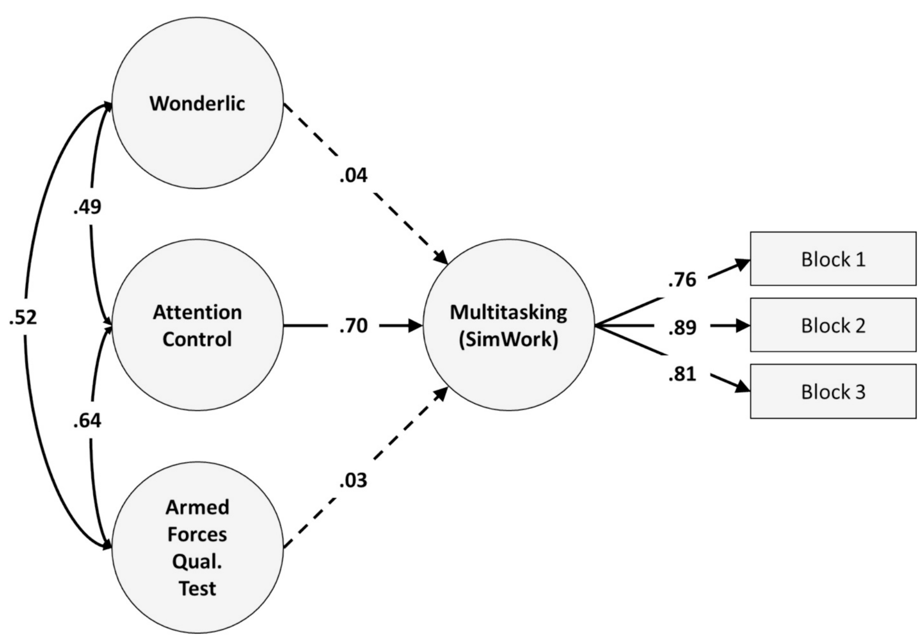

Fig. 5. Structural equation model with wonderlic, attention control, and armed forces qualification test latent factors predicting multitasking ability.
Solid paths are significant at $p < .05$; dashed paths are not significant. Predictor factor loadings are omitted for visual clarity but all were large and significant, $ps <$ 0.001. For attention control, the factor loadings were 0.72 for Stroop Squared, 0.75 for Flanker Squared, and 0.77 for Simon Squared. For the Wonderlic, the factor loadings were 0.90 for Parcel 1, 0.87 for Item Parcel 2, and 0.89 for Parcel 3. For the Armed Forces Qualification Test, the factor loadings were 0.48 for Word Knowledge, 0.57 for Paragraph Comprehension, 0.85 for Arithmetic Reasoning, and 0.76 for Mathematics Knowledge; thus, this factor was primarily driven by the numerical subtests. Model fit was acceptable, χ^2 (59) = 186.892, $p < .001$; CFI = 0.954, TLI = 0.939, RMSEA = 0.067, 90% CI [0.056, 0.078], SRMR = 0.044.

differences in processing speed may also figure into the strong association between attention control and simulated work performance in Study 1. In Study 2, we control for these other sources of variance and assess whether the relationship between attention control and simulated work multitasking remains robust.

Study 2 also included a complex test of psychomotor ability that the U.S. Navy, Marine Corps, and Coast Guard currently use as part of the Aviation Selection Test Battery for occupational classification purposes: the Performance Based Measures (Walker et al., 2007). In addition to a strong psychomotor component, the Performance Based Measures appears to demand a considerable amount of controlled attention. Consider its seven subtests: 1) Direction Orientation/Terrain Orientation 2) Dichotic Listening, 3) Two-dimensional Airplane Tracking, 4) Vertical Tracking, 5) combined Two-dimensional Airplane Tracking and Vertical Tracking, 6) combined Dichotic Listening, Two-dimensional Airplane Tracking, and Vertical Tracking, and 7) Emergency Scenarios (which includes two-handed tracking while performing emergency procedures). The dichotic listening test is a classic selective attention test in which participants must pay attention to one stream of auditory information while selectively ignoring another (Burgoyne et al., in press; Cherry, 1953; Conway, Cowan, & Bunting, 2001; Moray, 1959). The Two-dimensional Airplane Tracking and Vertical Tracking tests require selective attention to moving targets and crosshairs on the computer screen, as well as psychomotor ability to guide the crosshairs over the targets using hands-on-throttle-and-stick controllers. The Emergency Scenarios test has participants read lengthy instructions on how to use the throttle and joystick to mitigate three catastrophes (e.g., the engine catching on fire) and then implement those instructions under time constraints, all while performing vertical and two-dimensional tracking. Attention control has been linked to the ability to follow instructions and implement them motorically (Allen, Waterman, Yang, & Jaroslawska, 2022; Buszard et al., 2017; Engle, Carullo, & Collins, 1991; Jaroslawska, Gathercole, Logie, & Holmes, 2016).

The combined Performance Based Measures subtasks can be thought of as multitasks that compete for domain-general attentional resources. Performing the tasks concurrently likely runs up against an attentional bottleneck that constrains performance (Hambrick, Oswald, Darowski, Rench, & Brou, 2010). Importantly, many real-world "psychomotor" tasks share this multi-task structure, particularly early in learning. For instance, successfully piloting an aircraft depends on multi-limb coordination as well as attending to and correcting for changing features of the environment. Much of this demand is reduced as tasks become more proceduralized (Logan & Crump, 2011; Neisser, 1964; Spelke, Hirst, & Neisser, 1976).

Given these considerations, the primary "active ingredient" in the Performance Based Measures for predicting multitasking ability may be attention control. This would be useful to know from an applied perspective, since this could suggest augmentations or simplifications to the Performance Based Measures or other performance-based testing procedures. If our hypothesis is correct, by measuring attention control directly, we should be able to account for the validity of the Performance Based Measures for predicting multitasking ability. Nonetheless, several Performance Based Measures make strong psychomotor demands (e.g., manual tracking tasks) and, to the degree that such abilities affect multitasking, the Performance Based Measures should remain a significant predictor of performance even after accounting for attention control. For example, inefficient or wasteful movements could harm performance in both the unimanual and bimanual tracking tasks as well as our multitasking battery. Such common psychomotor demands of the Performance Based Measures could help explain the strong association between attention control and multitasking observed in Study 1. We investigated these possibilities using structural equation modeling and latent variable mediation analyses.

Our primary goal in Study 2 was to investigate the relative contribution of attention control, the Armed Forces Qualification Test, and the Performance Based Measures to performance in a series of simulated work multitasks while accounting for individual differences in working memory capacity and processing speed.

Our hypotheses were that 1) attention control's relationship with multitasking would not be eliminated by controlling for working memory capacity and processing speed, 2) that attention control would emerge as a dominant predictor of multitasking performance relative to the Armed Forces Qualification Test and Performance Based Measures, and that 3) attention control would mediate the relationship between the Performance Based Measures and multitasking performance, i.e., that the Performance Based Measures would predict simulated work multitasking only because of its relationship with attention control.

3.1. Method

3.1.1. Research ethics and consent

This study was approved by the Georgia Institute of Technology's Institutional Review Board under Protocol H20165. All participants provided written consent prior to participating in this study.

3.1.2. Participants

The study was conducted at the Georgia Institute of Technology (Georgia Tech) in Atlanta, Georgia, USA. All participants were fluent English speakers and between 18 and 35 years of age. We recruited participants from the Georgia Tech student body, other local colleges and universities, and the broader Atlanta community. Georgia Tech students enrolled in an undergraduate psychology course were given the option to receive 2.5 h of course credit or monetary compensation for each of the five sessions (see our Procedure for study 2). Participants from outside the university were recruited via message online message boards, fliers posted around the city (e.g., coffee shops, public transit stations, libraries), and word of mouth from other participants. The inclusion of participants from outside the university was intended to diversify the sample and improve the external validity of the study. A total of 305 participants completed all five sessions; 321 participants completed sessions one through four.

Demographic information is presented in Table 6. The age of the sample ranged from 18 to 35 ($M = 21.95$, $SD = 4.09$) and the majority had attended at least some college (86%). The sample was mostly female (58%), with males comprising 39% of the sample, and 1.6% identifying with another gender identity. Regarding the race/ethnicity of the participants, 41% selected Asian or Pacific Islander, 28% selected White, 14% selected Black or African American, and 17% selected "Other" or multiple categories. A small number of participants, 2.2%, provided no demographic information. The sample is thus large and racially/ethnically diverse, although smaller and more educated than that of Study 1, and with a larger percentage of female-identifying participants. Even so,

Table 6
Demographic information for Study 2.

Demographic	Statistic
Age	Mean: 22.0
	SD: 4.1
	Range: 18–35
	Missing: 2.2%
Gender:	Male: 38.6%
	Female: 57.6%
	Other/Self-Identify: 1.6%
	Missing: 2.2%
At least some college?	Yes: 86.0%
	No: 11.8%
	Missing: 2.2%
Race/Ethnicity	Black or African American: 13.1%
	Asian or Pacific Islander: 40.5%
	White: 27.7%
	Other/More than one category: 16.5%
	Missing: 2.2%

$N = 321$.

the sample reflects a broad range of demographic and sociocultural backgrounds.

3.1.3. Procedure

Participants were seated in individual testing rooms with a research assistant assigned to proctor each session. Research assistants were trained on the study protocol, including how to answer common participant questions and address minor technical difficulties. This was to ensure a uniform testing experience, to the degree possible. Further, participants read a common set of instructions for each task. The research assistant's job was to run each computerized cognitive test, ensure the participant understood the instructions, and make sure participants were following the rules of the lab, such as not using their phone during the study. Both the assistant and participant were masked due to COVID-19 restrictions. To further reduce contact, testing rooms were situated to minimize contact between the participant and the experimenter (e.g., a second mouse and keyboard were connected to the computer so that the experimenter could control the computer from a safe distance). All equipment was sanitized before and after each use.

Data were collected as part of a larger project, which consisted of >40 cognitive tasks administered over five sessions lasting 2.5 h each. We report on a subset of the data here, focusing specifically on the three-minute tests of attention control, the Armed Forces Qualification Test, and the Performance Based Measures. Further information regarding the scope of the data collection effort (except the Performance Based Measures subtest, which is not approved for public release) and other research products based on it can be found at the following link: https://osf.io/qbwem. For present purposes, the attention control measures and multitasking paradigms were administered during Sessions 1–4, and the Armed Forces Qualification Test and Performance Based Measures were administered during Session 5.

3.1.3.1. Armed forces qualification test. We administered the Armed Forces Qualification Test items used in Study 1.

3.1.3.1.1. Arithmetic reasoning. See Martin et al. (2020) and the Study 1 description.

3.1.3.1.2. Mathematical knowledge. See Martin et al. (2020) and the Study 1 description.

3.1.3.1.3. Word knowledge. See Martin et al. (2020) and the Study 1 description.

3.1.3.1.4. Paragraph comprehension. See Martin et al. (2020) and the Study 1 description.

3.1.3.2. Performance based measures subtest. The Performance Based Measures subtest from the Navy's Aviation Selection Test Battery consists of seven subtests: 1) Terrain Orientation, 2) Dichotic Listening, 3) Two-dimensional Airplane Tracking, 4) Vertical Tracking, 5) combined Two-dimensional Airplane Tracking and Vertical Tracking, 6) combined Dichotic Listening, Two-dimensional Airplane Tracking, and Vertical Tracking, and 7) Emergency Scenarios. Participants used the hands-on-throttle-and-stick controller (Fig. 6) for all subtests except for Terrain Orientation, for which they used the computer mouse. Five performance metrics are extracted from this task battery, described below. Because it is a proprietary test used by the United States Military, we are not at liberty to disclose how the Performance Based Measures is scored because the subtest weightings are not cleared for public release.

3.1.3.2.1. Terrain orientation. Participants determined the trajectory of an unmanned aerial vehicle based on the position of objects in the terrain below. On the left side of the screen, participants were shown a reference map that always displayed North at the top of the image. On the right side of the screen, participants were shown the "camera view" (i.e., bird's eye view) taken from their virtual unmanned aerial vehicle, which could be oriented in one of up to 12 directions relative to the reference map. The task was to indicate which direction the unmanned aerial vehicle was facing, given the camera view. The measure of performance was the number of correct responses during 24 trials. Importantly, this version of the task differed from the official spatial ability measure in the Performance Based Measures subtest, and is being considered as an upgrade from the official version.

Fig. 6. The hands-on-throttle-and-stick controllers used for the performance based measures.

3.1.3.2.2. Dichotic listening score. Participants were presented a series of numbers and letters via headphones to each ear and instructed to monitor a target ear while ignoring the other one. Their task was to press the trigger of the joystick when they heard an even number and to press the thumb button of the throttle when they heard an odd number. The measure of performance accounted for both the speed and accuracy of responses. Performance was measured during both the isolated dichotic listening test as well as during the combined Two-dimensional Airplane Tracking, Vertical Tracking, and dichotic listening task. See Walker et al. (2007) for additional details.

3.1.3.2.3. Two-dimensional airplane tracking score. Participants had to use the stick from a hands-on-throttle-and-stick controller to accurately move an aiming crosshair over a target airplane that changed direction and speed pseudo-randomly. The airplane could move any direction on the computer screen. The outcome measure reflects performance on the airplane tracking components of the single Two-dimensional Airplane Tracking task, the combined Two-dimensional Airplane Tracking and Vertical Tracking task, and the combined Dichotic Listening, Two-dimensional Airplane Tracking, and Vertical Tracking task. See Walker et al. (2007) for additional details.

3.1.3.2.4. Vertical tracking score. Participants had to use the throttle from a hands-on-throttle-and-stick controller to accurately move an aiming crosshair over a target airplane that changed direction and speed pseudo-randomly. The airplane could only move up or down (i.e., vertically) on the computer screen. The outcome measure reflects performance on the vertical tracking components of the isolated Vertical Tracking task, the combined Two-dimensional Airplane Tracking and vertical tracking task, and the combined Dichotic Listening, Two-dimensional Airplane Tracking, and Vertical Tracking task. See Walker et al. (2007) for additional details.

3.1.3.2.5. Emergency scenarios score. Participants were challenged to memorize and implement the appropriate motoric response to three emergency scenarios that occurred while performing the simultaneous Airplane Tracking and vertical tracking tasks. All three emergency scenarios (e.g., fire light, engine light, and propeller light) required multiple motor responses using the hands-on-throttle-and-stick controller (e.g., adjust fuel to low, power to low, and reset the emergency scenario). The measure of performance reflects the accuracy and speed of participants' responses to these emergency scenarios while performing Vertical and Two-dimensional Tracking tasks. See Walker et al. (2007)

for additional details.

3.1.3.3. Attention control. We administered the same battery of three-minute attention control tests used in Study 1.

3.1.3.3.1. Stroop Squared. See Study 1.

3.1.3.3.2. Flanker Squared. See Study 1.

3.1.3.3.3. Simon Squared. See Study 1.

3.1.3.4. Working memory capacity. We administered two complex span working memory tasks.

3.1.3.4.1. Advanced symmetry span (Draheim et al., 2018). Participants attempted to remember a series of spatial locations in a 4 × 4 matrix. Each spatial memorandum was interleaved with a processing task in which participants judged whether a 16 × 16 configuration of black and withe squares was symmetrical about the vertical midline. On each trial, participants are presented with a symmetry judgement, followed by a 4 × 4 grid with one square highlighted in red. The location of the red square was the to-be-remembered spatial location. Participants completed a variable number of alternations (2–7) until a recall screen appeared. Participants then attempted to recall the locations of the red square in their correct serial order. There was a total of 12 trials (2 blocks of 6 trials), set-sizes ranged from 2 to 7, and each set-size occurred twice (once in each block). The dependent variable is the edit distance score (see Gonthier, 2023).

3.1.3.4.2. Advanced rotation span (Draheim et al., 2018). Participants tried to remember a series of directional arrows of varying sizes. These were interleaved with a mental rotation task in which participants mentally rotated a letter and decided whether it is mirror reversed. On each trial, participants first solved a mental rotation problem followed by the presentation of a single arrow with a specific direction (8 possible directions; the four cardinal and four ordinal directions) and specific size (small or large). Both the direction and size of the arrow were the to-be-remembered features. This alternation continued until a variable set-size of arrows was presented, when participants tried to recall the set in their correct serial position. There are 12 trials (2 blocks of 6 trials), set-sizes ranged from 2 to 7, and each set- size occurs twice (once in each block). Once again, the dependent variable is the edit distance score (see Gonthier, 2023).

3.1.3.5. Processing speed. We administered three computerized processing speed tasks.

3.1.3.5.1. Digit comparison (Draheim et al., 2021; 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 30s 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.

3.1.3.5.2. Letter comparison (Draheim et al., 2021; 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 three, six, or nine letters.

3.1.3.5.3. Pattern comparison (Redick et al., 2012; 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 30s of trials and

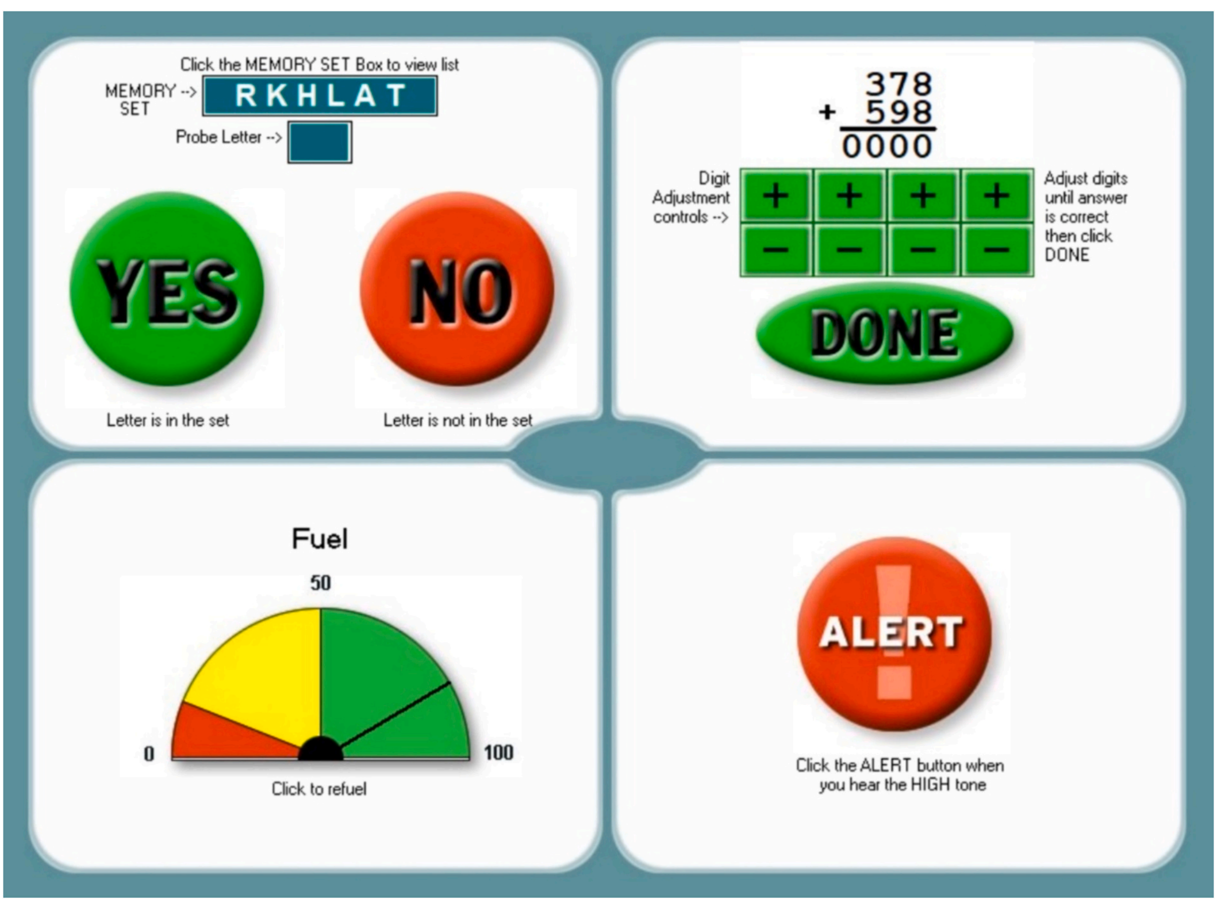

Fig. 7. The Synthetic Work (SynWin) interface.
The four subtasks are: Memory Search (top-left); Math (top-right); Visual Monitoring (bottom-left); and Auditory Monitoring (bottom-right).

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.

3.1.3.6. Multitasks. We administered a battery of three simulated work multitasks.

3.1.3.6.1. Synthetic Work for Windows (SynWin; Elsmore, 1994; Fig. 7). 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. The outcome measure was the average score across three five-minute test blocks. Task details are presented in Martin et al. (2020).

3.1.3.6.2. FosterMT (Burgoyne, Tsukahara, Mashburn, Pak, & Engle, 2023; Martin et al., 2020; Figure 8). The four subtasks included simple mathematics, word recall, and two visual monitoring subtasks. One visual monitoring subtask required participants to monitor a battery; allowing the battery to die or to charge past full capacity resulted in lost points, and participants must intermittently either "plug-in" or "unplug" the battery to avoid either scenario. The second visual monitoring task required participants to click on an onscreen disk when it starts to spin. The outcome measure was the average score across three five-minute test blocks.

3.1.3.6.3. Control tower (Redick et al., 2016; Fig. 9). Participants were given a primary task and multiple distractor tasks to complete over one ten-minute block. The primary task entailed a symbol substitution task involving numbers, letters, and symbols according to a set of rules. 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. Task details are provided in Martin et al. (2020).

3.1.4. 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. We set scores that fell below chance performance (i.e., performance which was below what would be expected if participants were responding based on guessing alone) to missing. For the attention control tasks, this meant removing any score which fell below zero. For other tasks, such as the Terrain Orientation Task, chance performance meant dividing the proportion of total responses (i.e., 1.0) by the total number of response options (i.e., 12; see Table 7). This resulted in the removal of 22 scores on the Flanker Squared task and 18 cases of the Terrain Orientation Task. We also screened for outliers, which we define as any value with a standardized score (i.e., z-score) ±3.5 standard deviations from the sample mean. After setting these outlying scores to missing, we recomputed standard deviations and re-checked for outliers. We repeated this process until no further outliers were identified (see Table 7). Additionally, because the Armed Forces Qualification Test is a composite variable based on four subtests, a missing score on one subtest would render the composite uninterpretable. Therefore, if a participant had an outlying score on any one subtest, we set their other Armed Forces Qualification Test scores to missing, resulting in a final sample of 291 cases for all Armed Forces Qualifying Test variables.

3.1.5. Modeling approach and fit statistics

Modeling details are the same as Study 1.

3.2. Results

3.2.1. Descriptive statistics and reliability

Descriptive statistics and reliability estimates are found in Table 8. The split-half reliability of the attention control tests ranged from 0.94 to 0.97. For the multitasks, the three blocks of SynWin had a Cronbach's alpha of 0.90, and the three blocks of FosterMT had a Cronbach's alpha of 0.95. Cronbach's alpha for the subtests of the Armed Forces Qualification Test ranged from 0.45 (Paragraph Comprehension) to 0.77 (Arithmetic Reasoning). For the Performance Based Measures, Cronbach's alpha could only be computed for the Terrain Orientation task (0.87). For working memory capacity, symmetry span had a Cronbach's

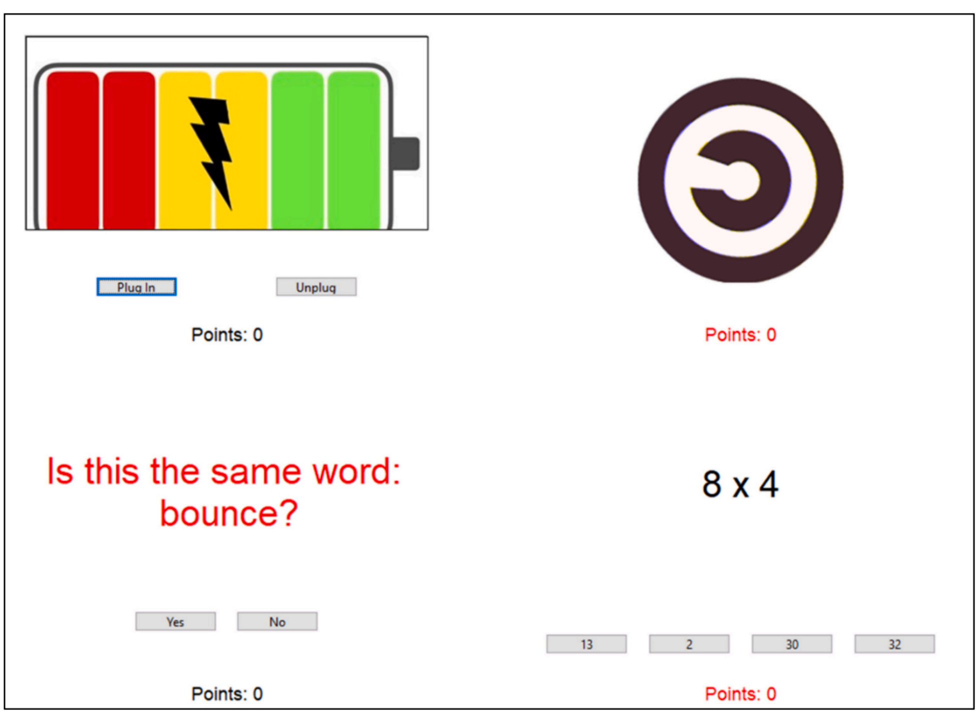

Fig. 8. The FosterMT interface.
The four subtasks are: Visual Monitoring – Battery (top-left); Visual Monitoring – Disk (top-right); Word Recall (bottom-left); and Math Problems (bottom-right).

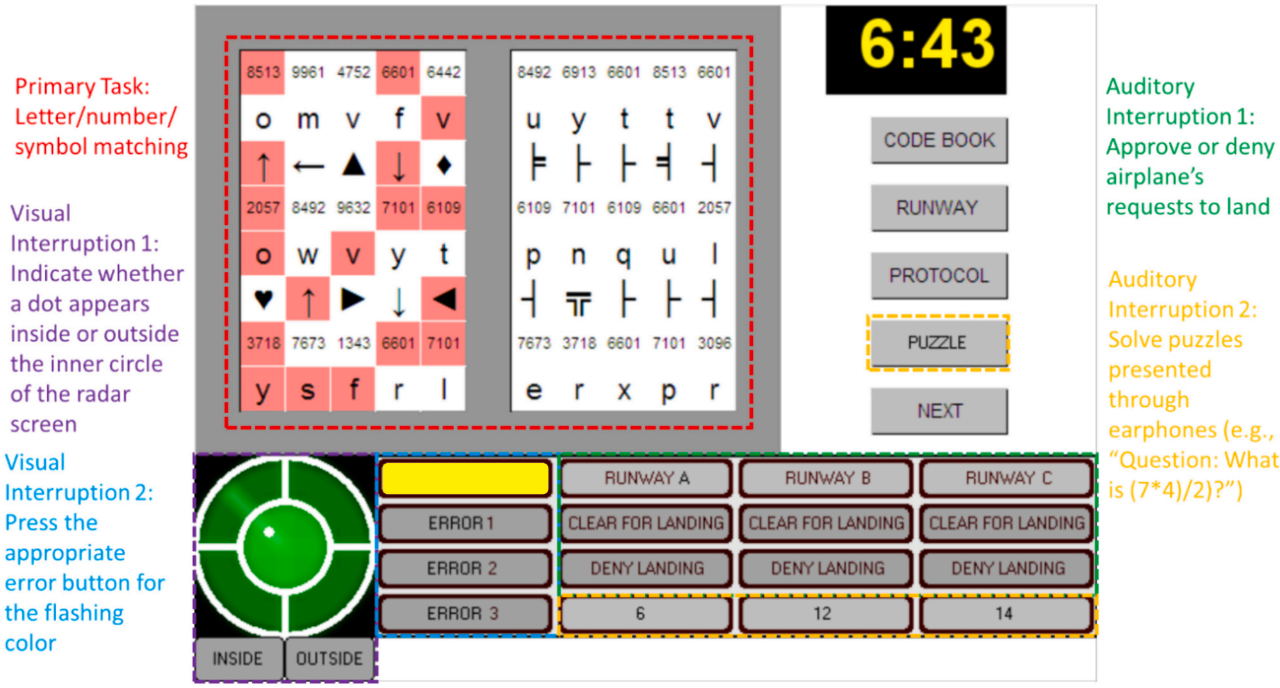

Fig. 9. Labeled snapshot of the control tower interface.
The font color of the descriptions correspond to the element overlaid colored squares.

Table 7
Summary of removed cases for Study 2.

Criterion	Task	Cases Removed	
Chance Performance			*Chance Cutoff*
	Flanker Squared	22	0 points
	Terrain Orientation Task	18	8.3%
Outlier (±3.5 SDs)			*Outlier Passes (#)*
	Mathematics Knowledge	1	1
	Word Knowledge	2	1
	Control Tower (D)	10	2
	Paragraph Comprehension	6	2
	Control Tower (D)	10	2
	SynWin	9	3
	PBM Airplane Tracking Score	4	2
	PBM Vertical Tracking Score	2	1
	Rotation Span	3	1
	Symmetry Span	1	1

alpha of 0.76, while rotation span had a Cronbach's alpha of 0.73. Finally, the three processing speed tasks had Spearman-Brown corrected split-half reliability estimates ranging from 0.82 to 0.94.

Table 8 reveals several notable trends. The most important is that mean performance on the Armed Forces Qualification Test and attention control tasks is higher in Study 2 than in Study 1 (see Table 3). This is notable given concerns about participants in Study 1 looking up the answers to the knowledge tests because those tasks were administered online. While this does not rule out cheating in Study 1 per se, this pattern may signal ability differences across the two samples. We discuss this possibility further in our discussion.

3.2.2. *Bivariate correlations*

Task-level correlations were computed using pairwise deletion and are presented in Table 9; all were expected to be significant and positive in direction. The predictor measures were significantly correlated with the four indicators of multitasking performance (e.g., SynWin, FosterMT, Control Tower-Primary, Control Tower-Distractor). For instance, the three Squared tests of attention control had an average correlation of $r = 0.43$ with the multitasking measures. The other predictors had average correlations with multitasking equal to $r = 0.33$ for the Armed Forces Qualification Test subtests, $r = 0.29$ for the Performance Based Measures, $r = 0.25$ for working memory capacity, and $r = 0.42$ for processing speed. Contrary to our expectations, there were several non-significant correlations between the Performance Based Measures subtests and measures of working memory capacity and processing speed, and the Word Knowledge subtest had several non-significant correlations with other measures. We next used structural equation modeling to clarify the latent structure of the correlation matrix.

3.2.3. *Structural equation modeling*

We examined the relationships between latent variables in four steps. First, we estimated the correlations between factors using confirmatory factor analysis. Next, we evaluated whether attention control predicted multitasking performance after controlling for working memory and processing speed. This step established that the attention control tasks measure more than mere working memory capacity and/or processing speed. We then carried forward the non-redundant predictors from the second model to a third structural equation model, which added the Armed Forces Qualification Test and Performance Based Measures as correlated predictors of multitasking performance. This allowed us to test the unique contribution of each latent variable to simulated work performance. Finally, we tested whether attention control, the Armed Forces Qualification Test, and processing speed mediated the relationship between the Performance Based Measures and multitasking ability.

Our confirmatory factor analysis began by specifying all observed measures as indicators of their respective latent variables. These latent variables were then allowed to freely correlate. Results of the confirmatory factor analysis are summarized in Table 10; model fit was satisfactory, $\chi^2(174) = 332.703$, $p < .001$; CFI $= 0.925$, TLI $= 0.909$, RMSEA $= 0.053$, 90% CI [0.045, 0.062], SRMR $= 0.057$. All latent

Table 8
Descriptive statistics for Study 2.

Measure	*N*	*M*	*SD*	Skew	Kurtosis	Reliability
SynWin	306	3257.63	544.02	−0.30	0.48	0.89[α]
FosterMT	302	96,011.66	26,343.41	−0.20	0.05	0.95[α]
Control Tower (P)	316	101.59	31.28	−0.10	0.37	–
Control Tower (D)	306	25.83	2.45	−0.95	0.46	–
Flanker Squared	299	40.13	14.16	0.13	−0.22	0.97[b]
Stroop Squared	321	41.30	12.66	−0.08	0.17	0.94[b]
Simon Squared	320	67.82	9.44	−0.19	0.14	0.94[b]
AFQT Composite[c]	290	91.71	11.24	−0.36	−0.08	0.83[α]
AFQT Arithmetic Reasoning	290	16.39	4.03	−0.49	−0.24	0.77[α]
AFQT Math Knowledge	290	14.90	2.64	−0.91	0.07	0.71[α]
AFQT Word Knowledge	290	20.13	3.00	−0.23	0.38	0.64[α]
AFQT Paragraph Comprehension	290	10.08	1.53	−1.09	0.91	0.45[α]
PBM Terrain Orientation	139	0.44	0.24	0.30	−1.12	0.87[α]
PBM Dichotic Listening Score	259	24.84	10.55	−0.54	−0.62	–
PBM Airplane Tracking Score	255	7.31	7.62	1.21	1.12	–
PBM Vertical Tracking Score	257	25.05	10.04	0.14	0.04	–
PBM Emergencies Score	259	9.94	10.14	0.01	−1.69	–
Symmetry Span	310	34.97	8.83	−0.47	−0.15	0.76[α]
Rotation Span	310	29.45	8.48	−0.39	0.05	0.73[α]
Digit Comparison	307	29.90	5.51	−0.45	0.02	0.88[b]
Letter Comparison	307	20.53	4.10	0.12	0.39	0.82[b]
Pattern Comparison	306	39.06	6.01	−0.09	−0.22	0.94[b]

AFQT = Armed Forces Qualification Test. PBM = Performance Based Measures.
— = Reliability could not be computed. Control Tower (P) = Primary, Control Tower (D) = Distractor.
[α] Cronbach's alpha.
[b] Split-half reliability with Spearman-Brown correction.
[c] Cronbach's alpha for the AFQT was computed using participants' accuracy data across the entire set of items with no adjustment for subtest weighting.

correlations, with the exception of those between the Armed Forces Qualification Test and working memory capacity and processing speed, were moderate to very large. In particular, the correlations between multitasking and attention control/processing speed are exceedingly large at $r = 0.91$ and $r = 0.89$, respectively. Before continuing, we must establish whether these constructs are dissociable. We assessed this using likelihood ratio tests, collapsing across (i.e., combining) the factors in question and evaluating the reduction in model fit. Despite being strongly correlated, attention control and multitasking could not be combined into a unitary factor without significantly reducing model fit, $\Delta\chi^2(5) = 26.630$, $p < .001$. Thus, despite being very strongly related, attention control and multitasking factors were statistically distinguishable. The same was true of processing speed: collapsing processing speed and multitasking onto one factor significantly reduced model fit, $\Delta\chi^2(5) = 46.82$, $p < .001$. Having established that the factors are isolable, we proceeded to our next analysis.[2]

Our next model assessed the relationships among multitasking, attention control, working memory capacity, and processing speed. We sought to establish whether relationships remained among the other constructs once variance related to processing speed was removed. To this end, we specified a structural equation model with attention control and working memory capacity as correlated predictors of multitasking. We also specified processing speed as a predictor of the other three latent variables, such that any structural relationships between attention control, working memory capacity, and multitasking would be independent of processing speed. The model is depicted in Fig. 10; model fit was acceptable, $\chi^2(48) = 99.289$, $p < .001$, CFI = 0.968, TLI = 0.951, RMSEA = 0.056, 90% CI [0.042, 0.074], SRMR = 0.044. Processing speed reliably predicted individual differences in attention control, working memory capacity, and multitasking. Though weaker than our earlier confirmatory factor analysis (see Table 10), the residual correlation between attention control and working memory capacity after controlling for processing speed remained significant, $r = 0.32$, $p = .007$. Most importantly, attention control reliably predicted multitasking even accounting for processing speed and working memory capacity, $\beta = 0.52$, $p < .001$. No unique relationship remained between working memory capacity and multitasking, $\beta = 0.06$, $p = .34$. Since it bore no unique relationship with multitasking in this model, we do not carry working memory capacity forward as a predictor in the following analyses.

We next specified a model with the remaining latent variables, including the Armed Forces Qualification Test and Performance Based Measures, as correlated predictors of multitasking. Shown in Fig. 11, this model fit acceptably, $\chi^2(142) = 285.974$, $p < .001$; CFI = 0.925, TLI = 0.910, RMSEA = 0.056, 90% CI [0.047, 0.066], SRMR = 0.057. Attention control ($\beta = 0.31$, $p = .045$), the Armed Forces Qualification Test ($\beta = 0.36$, $p < .001$), and processing speed ($\beta = 0.45$, $p < .001$) each significantly predicted multitasking ability, whereas the Performance Based Measures did not ($\beta = 0.02$, $p = .86$). The significant predictive relationship between the Armed Forces Qualification Test and multitasking in these data may affirm our concern about the un-proctored testing environment from Study 1; we consider this further in our Discussion. The non-significant predictive relationship between the Performance Based Measures and multitasking suggests that the earlier correlation between the Performance Based Measures and multitasking ability (see Table 10; $r = 0.79$, $p < .001$) mainly reflects variance that is also shared with the other predictors. We investigated this further using mediation analyses.

We next tested three models in which attention control, the Armed Forces Qualification Test, or processing speed mediated the relationship between the Performance Based Measures and multitasking ability. The fit of all three models was acceptable (see Table 11). As shown in the top panel of Fig. 12, the effect of the Performance Based Measures on multitasking performance could be completely explained by its relationship with attention control. After accounting for attention control, the Performance Based Measures was no longer a direct predictor of multitasking ($\beta = 0.15$, $p = .27$). Meanwhile, both the direct effect of attention control ($\beta = 0.79$, $p < .001$) and the indirect effect of the

[2] JASP does not report 95% confidence intervals for standardized structural equation model estimates, so we replicated this analysis in R using the *lavaan* package. The 95% confidence interval for the latent correlations between multitasking and attention control (0.85–0.97) and between multitasking and processing speed (0.82–0.95) did not contain 1.00. This corroborates the results of the likelihood ratio tests.

Table 9
Correlations between major variables of Study 2.

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.
1. SynWin	–																				
2. FosterMT	0.61	–																			
3. Control Tower (P)	0.51	0.59	–																		
4. Control Tower (D)	0.31	0.34	0.23	–																	
5. Flanker Squared	0.49	0.55	0.37	0.33	–																
6. Stroop Squared	0.42	0.42	0.40	0.34	0.48	–															
7. Simon Squared	0.49	0.65	0.50	0.24	0.52	0.52	–														
8. AFQT Composite	0.48	0.45	0.44	0.35	0.37	0.27	0.31	–													
9. AFQT Arithmetic Reasoning	0.45	0.44	0.37	0.30	0.41	0.26	0.27	0.74	–												
10. AFQT Math Knowledge	0.42	0.43	0.37	0.32	0.47	0.33	0.27	0.64	0.60	–											
11. AFQT Word Knowledge	0.25	0.17	0.25	0.18	0.11	0.09	0.13	0.78	0.28	0.22	–										
12. AFQT Paragraph Comprehension	0.31	0.35	0.34	0.27	0.19	0.20	0.29	0.63	0.32	0.28	0.34	–									
13. PBM Terrain Orientation	0.43	0.39	0.34	0.29	0.42	0.42	0.34	0.37	0.45	0.49	−0.01	0.30	–								
14. PBM Dichotic Listening Score	0.38	0.40	0.31	0.24	0.34	0.22	0.34	0.26	0.13	0.20	0.18	0.29	0.27	–							
15. PBM Airplane Tracking Score	0.28	0.28	0.19	0.23	0.29	0.17	0.36	0.23	0.18	0.22	0.11	0.24	0.39	0.26	–						
16. PBM Vertical Tracking Score	0.27	0.30	0.28	0.23	0.35	0.24	0.29	0.29	0.28	0.29	0.11	0.22	0.19	0.31	0.23	–					
17. PBM Emergencies Score	0.24	0.20	0.27	0.26	0.28	0.25	0.22	0.30	0.20	0.27	0.19	0.22	0.28	0.42	0.19	0.29	–				
18. Symmetry Span	0.37	0.35	0.26	0.15	0.32	0.29	0.31	0.26	0.30	0.35	0.05	0.17	0.36	0.23	0.16	0.31	0.22	–			
19. Rotation Span	0.29	0.23	0.17	0.16	0.27	0.27	0.24	0.11	0.19	0.26	−0.05	0.04	0.16	0.19	0.10	0.25	0.14	0.53	–		
20. Digit Comparison	0.54	0.58	0.50	0.22	0.38	0.36	0.55	0.29	0.27	0.30	0.11	0.22	0.27	0.30	0.20	0.24	0.12	0.25	0.21	–	
21. Letter Comparison	0.41	0.47	0.44	0.17	0.28	0.30	0.42	0.21	0.16	0.26	0.10	0.14	0.22	0.20	0.07	0.23	0.16	0.22	0.18	0.61	–
22. Pattern Comparison	0.47	0.50	0.45	0.32	0.40	0.42	0.50	0.33	0.28	0.27	0.15	0.28	0.35	0.29	0.15	0.27	0.23	0.34	0.30	0.49	0.41

Boldface = $p < .05$. Pairwise N ranges from 243 to 320 for everything except correlations with PBM: Terrain Orientation (pairwise N ranged from 126 to 139 for those correlations). AFQT = Armed Forces Qualification Test, PBM = Performance Based Measures, Control Tower (P) = Primary, Control Tower (D) = Distractor.

Table 10
Correlations between factors in Study 2.

Factor	1.	2.	3.	4.	5.
1. Multitasking Ability	–				
2. Attention Control	**0.91**	–			
3. Armed Forces Qualification Test	**0.62**	**0.45**	–		
4. Performance Based Measures	**0.79**	**0.80**	**0.56**	–	
5. Working Memory Capacity	**0.52**	**0.52**	**0.27**	**0.58**	–
6. Processing Speed	**0.89**	**0.80**	**0.37**	**0.59**	**0.42**

Bold values are statistically significant at $p < .001$. All factor loadings were large and statistically significant, $ps < 0.001$. For the multitasking factor, loadings ranged from 0.45 (Control Tower-D) to 0.83 (FosterMT). The factor loadings for attention control ranged from 0.64 (Stroop Squared) to 0.79 (Simon Squared). For the Performance Based Measures, loadings ranged from 0.44 (Two-dimensional Airplane Tracking Score) to 0.68 (Terrain Orientation). For the Armed Forces Qualification Test, loadings ranged from 0.37 (Word Knowledge) to 0.77 (Arithmetic Reasoning). Symmetry span had a loading of 0.85 onto the working memory capacity factor, and rotation span had a loading of 0.62. Finally, loadings for processing speed ranged from 0.67 (Pattern Comparison) to 0.81 (Digit Comparison). Model fit was acceptable, χ^2 (174) = 332.703, $p < .001$; CFI = 0.925, TLI = 0.909, RMSEA = 0.053, 90% CI [0.045, 0.062], SRMR = 0.057.

Performance Based Measures through attention control (β = 0.63, Wald Z = 4.96, $p < .001$) were substantial and statistically significant.[3] Thus, attention control on its own was sufficient to account for the Performance Based Measure's prediction of multitasking ability.

By comparison, the Armed Forces Qualification Test and processing speed only partially mediated the effect of the Performance Based Measures on multitasking ability (see middle and bottom panels of Fig. 12). This is indicated by a significant direct effect of the Performance Based Measures on multitasking after accounting for the Armed Forces Qualification Test (β = 0.47, $p < .001$) and processing speed (β = 0.40, $p < .001$). Both the indirect effect of the Performance Based Measures through the Armed Forces Qualification Test (β = 0.32, Wald Z = 4.53, $p < .001$) and through processing speed (β = 0.38, Wald Z = 5.94, $p < .001$) were also significant. Thus, the Armed Forces Qualification Test and processing speed could reduce but not eliminate the relationship between the Performance Based Measures and multitasking ability, whereas attention control fully explained the relationship.[4]

Given that attention control fully mediated the relationship between latent factors representing the Performance Based Measures and multitasking ability, we also investigated whether full mediation occurred at the subtask-level, testing each of the five subtask measures from the Performance Based Measures (i.e., the Terrain Orientation Test, Dichotic Listening, Two-dimensional Airplane Tracking, Vertical Tracking, and Emergency Scenarios), one at a time. The fit statistics for each model are displayed in Table 12; model fit was acceptable in each case. (See Table 12.)

As shown in Fig. 13, attention control fully mediated the relationship between four of the five measures from the Performance Based Measures and multitasking ability and partially mediated the fifth. Only one of the Performance Based Measures, Dichotic Listening, significantly predicted multitasking ability after accounting for attention control, and its direct effect was relatively small (β = 0.12, $p = .031$). The indirect path from

[3] As a robustness check, we also ran a model with the Performance Based Measures specified as the mediator of attention control's effect on multitasking; the direct effect of attention control was significant (β = 0.79, $p < .001$), but the indirect effect through the Performance Based Measures was not (β = 0.12, Wald Z = 1.14, $p = .32$). This indicates that the Performance Based Measures did not mediate the attention control-multitasking relationship.

[4] As a robustness check, the Performance Based Measures partially mediated the effect of the Armed Forces Qualification Test on multitasking; both the direct effect (β = 0.46, $p < .001$) and indirect effect (β = 0.33, Wald Z = 4.10, $p < .001$) were significant. This pattern repeated for processing speed (direct effect: β = 0.65, $p < .001$; indirect effect: β = 0.23, Wald Z = 4.91, $p < .001$).

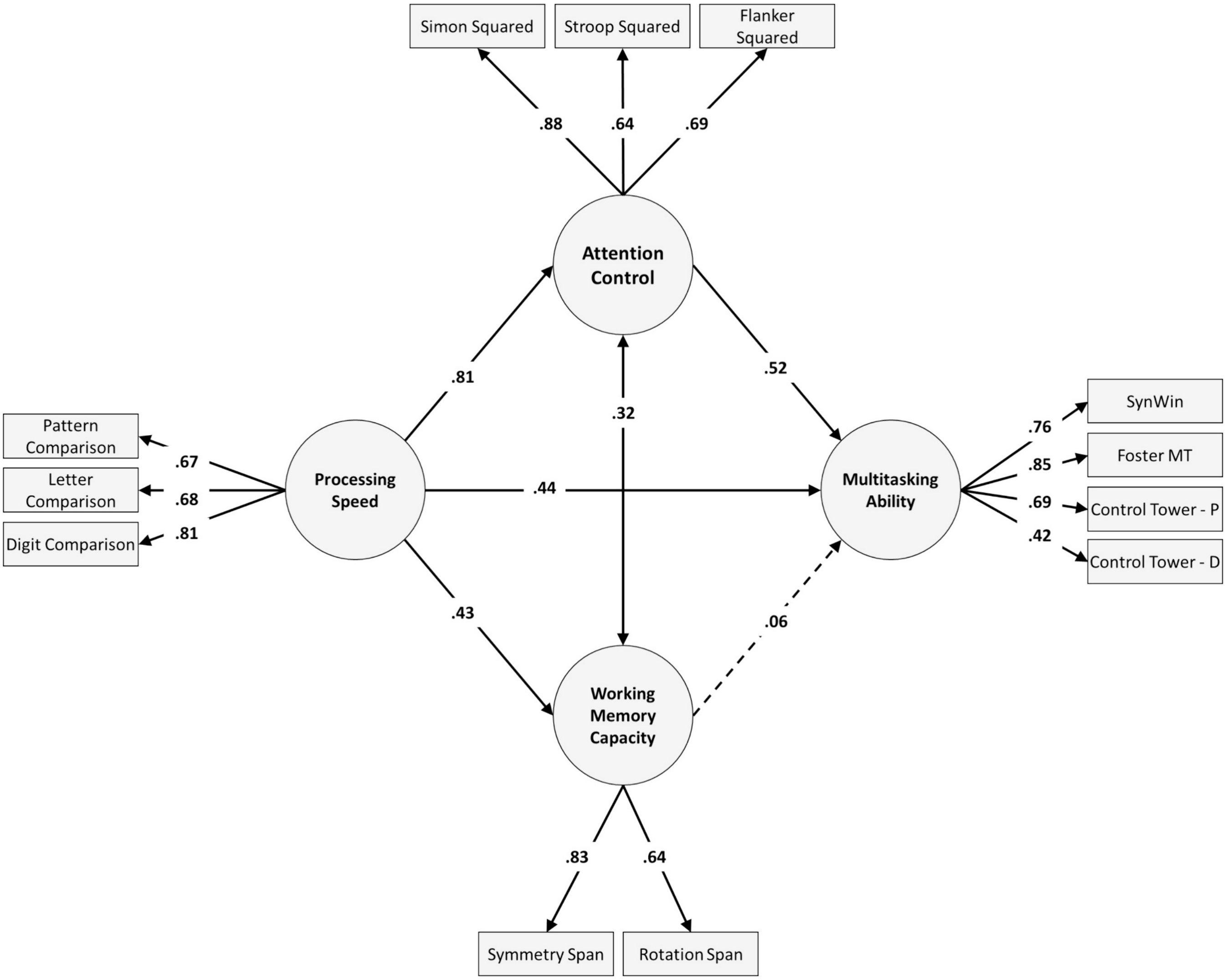

Fig. 10. Relationships among attention control, working memory capacity, and multitasking after accounting for processing speed.
Solid paths are significant at $p < .05$; dashed paths are not statistically significant. Model fit was acceptable, $\chi^2(48) = 99.289$, $p < .001$; CFI = 0.964, TLI = 0.951, RMSEA = 0.056, 90% CI [0.042, 0.074], SRMR = 0.044.

the Performance Based Measures through attention control to multitasking ability was statistically significant and substantial for all five measures. Thus, the validity of the Performance Based Measures for predicting multitasking performance appears largely attributable to attention control. By comparison, the Armed Forces Qualification Test failed to fully mediate the relationship between two Performance Based Measure subtests and multitasking (Dichotic Listening direct effect: $\beta = 0.32$, $p < .001$; Two-handed Airplane Tracking direct effect: $\beta = 0.16$, $p = .005$). Processing speed failed to fully mediate any of the relationships between the Performance Based Measures subtests and multitasking (the smallest direct effect was for the Vertical Tracking Score, $\beta = 0.11$, $p = .058$).

3.2.4. Gender differences

In our final set of exploratory analyses, we compared gender differences in performance on the attention control measures, the Armed Forces Qualification Test, the Performance Based Measures, and processing speed.[5] This analysis was motivated by the large gender differences found for psychomotor ability in the Air Force's Test of Basic Aviation Skills (Trent & Aguilar, 2020). The Performance Based Measures uses a vertical tracking task whereas the Test of Basic Aviation Skills uses a horizontal tracking task. Both assessments also include a two-dimensional tracking task, which is highly similar across the two batteries. Thus, it is likely that the Performance Based Measures has similar gender-based differences (Fatolitis et al., 2010).

If attention control measures not only explain the validity of the Performance Based Measures for predicting multitasking ability, but also show smaller differences in performance between men and women, then they may be preferable from a measurement and selection perspective. Importantly, the focus of these analyses is *not* on the absolute magnitude of subgroup differences, but on the *relative* size of the subgroup differences *across measures*. The absolute magnitude of subgroup differences is uninformative because the sample is not nationally representative. Selection effects (e.g., differential study recruitment and/or participation) can drive the absolute magnitude of subgroup differences up or down. Thus, the overall magnitude of subgroup differences should be seen as an artifact of participant sampling.

Table 13 provides the standardized differences in performance on the ability measures, comparing males and females based on Welch's *t*-tests. We report Cohen's *d* values and their respective 95% confidence

[5] We did not have a sufficient sample size to investigate subgroup differences in performance based on race/ethnicity.

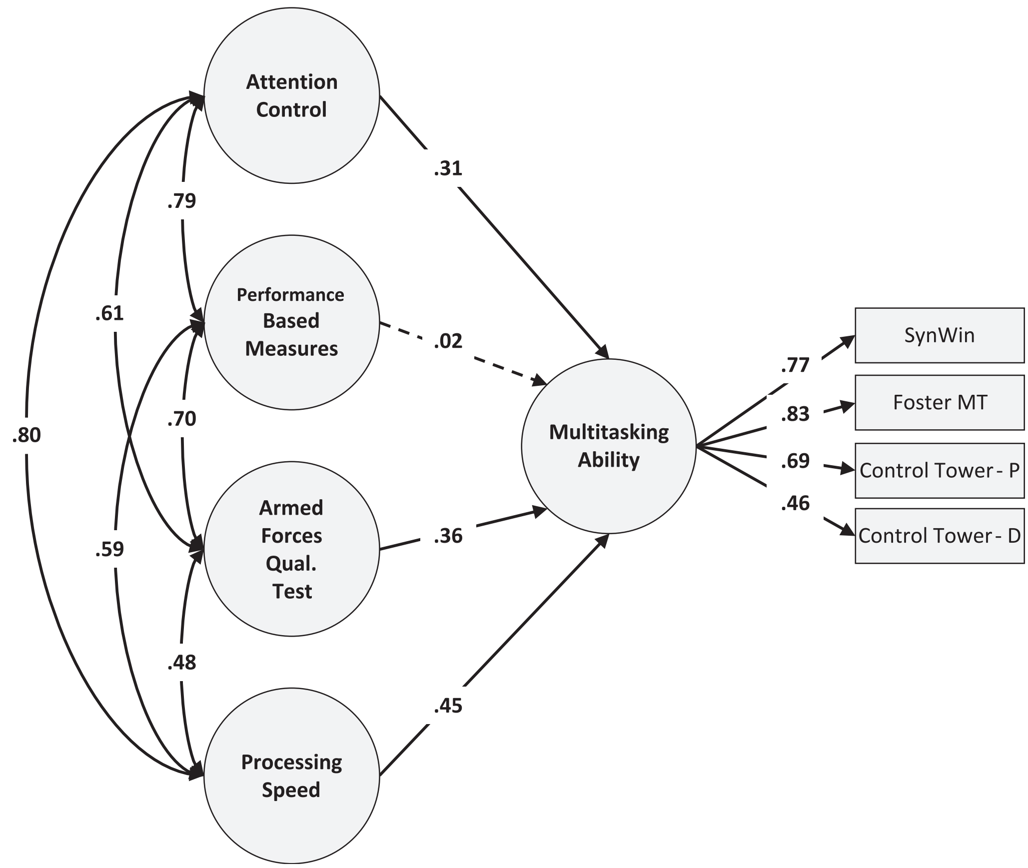

Fig. 11. Structural equation model with attention control, AFQT, PBM, and processing speed latent factors predicting multitasking ability.
Solid paths are significant at $p < .05$; dashed paths are not statistically significant. Predictor factor loadings are omitted for visual clarity but all were large and significant, *p*s < 0.001. For attention control, these ranged from 0.64 for Stroop Squared to 0.80 for Simon Squared. For the Performance Based Measures, loadings ranged from 0.45 (Two-dimensional Airplane Tracking Score) to 0.67 (Terrain Orientation). For the Armed Forces Qualification Test, loadings ranged from 0.38 (Word Knowledge) to 0.77 (Arithmetic Reasoning); thus, this latent factor was primarily driven by the mathematics subtests. Loadings for processing speed ranged from 0.67 (Pattern Comparison) to 0.81 (Digit Comparison). Model fit was acceptable, χ^2 (142) = 285.974, $p < .001$; CFI = 0.925, TLI = 0.910, RMSEA = 0.056, 90% CI [0.047, 0.066], SRMR = 0.057.

Table 11
Fit indices for models with attention control, AFQT, and processing speed mediating the relationship between each performance based measures and multitasking performance.

Mediator	χ^2(df), *p*-value	CFI	TLI	RMSEA [90% CI]	SRMR
Attention control	105.18(51), $p < .001$	0.949	0.934	0.058[0.042,0.073]	0.049
Armed Forces Qualification Test	124.71(62) $p < .001$	0.932	0.915	0.056[0.042,0.070]	0.056
Processing Speed	78.346(51), $p < .001$	0.973	0.965	0.041[0.021,0.058]	0.049

CFI = Confirmatory Fit Index. TLI = Tucker-Lewis Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual.

intervals. This information is redundant to the corresponding *t*-test, since any Cohen's *d* with a confidence interval containing zero corresponds to a non-significant effect of gender. However, evaluating effect sizes rather than *t*-values makes it easier to evaluate the magnitude of gender effects across the different measures. In these models, a negative Cohen's *d* indicates that those identifying as men scored higher than those identifying as women.

Three of the five Performance Based Measures demonstrated statistically significant differences between men and women (Cohen's *d*s of −0.51 [95% CI: −0.88, −0.15] for Terrain Orientation, −0.70 [95% CI: −0.96, −0.43] for Vertical Tracking, and − 1.06 [95% CI: −1.35, −0.77] for Two-dimensional Airplane Tracking); the remaining two measures had Cohen's *d*s of −0.09 [95% CI: −0.35, 0.16] for Dichotic Listening and − 0.17 [95% CI: −0.43, 0.09] for Emergency Scenarios. The average group difference across the five Performance Based Measures was $d =$ −0.51, favoring men. By comparison, the three Squared tests of attention control revealed gender differences of $d = -0.13$ [95% CI: −0.36, 0.10] for Stroop Squared, $d = -0.46$ [95% CI: −0.72, −0.22] for Flanker Squared, and $d = -0.49$ [95% CI: −0.70, −0.22] for Simon Squared, with an average group difference of $d = -0.36$. The subtests of the Armed Forces Qualification Test yielded gender differences ranging from $d = -0.06$ [95% CI: −0.31, 0.18] for paragraph comprehension to $d = -0.76$ [95% CI: −1.01, −0.51] for arithmetic reasoning. The Armed Forces Qualification Test composite had a group difference of $d = -0.47$. Finally, none of the processing speed measures showed any significant differences between men and women, having an average group difference of $d = -0.12$.

Thus, in addition to statistically explaining the relationship between the Performance Based Measures and simulated work multitasking, tests of attention control may demonstrate relatively small gender

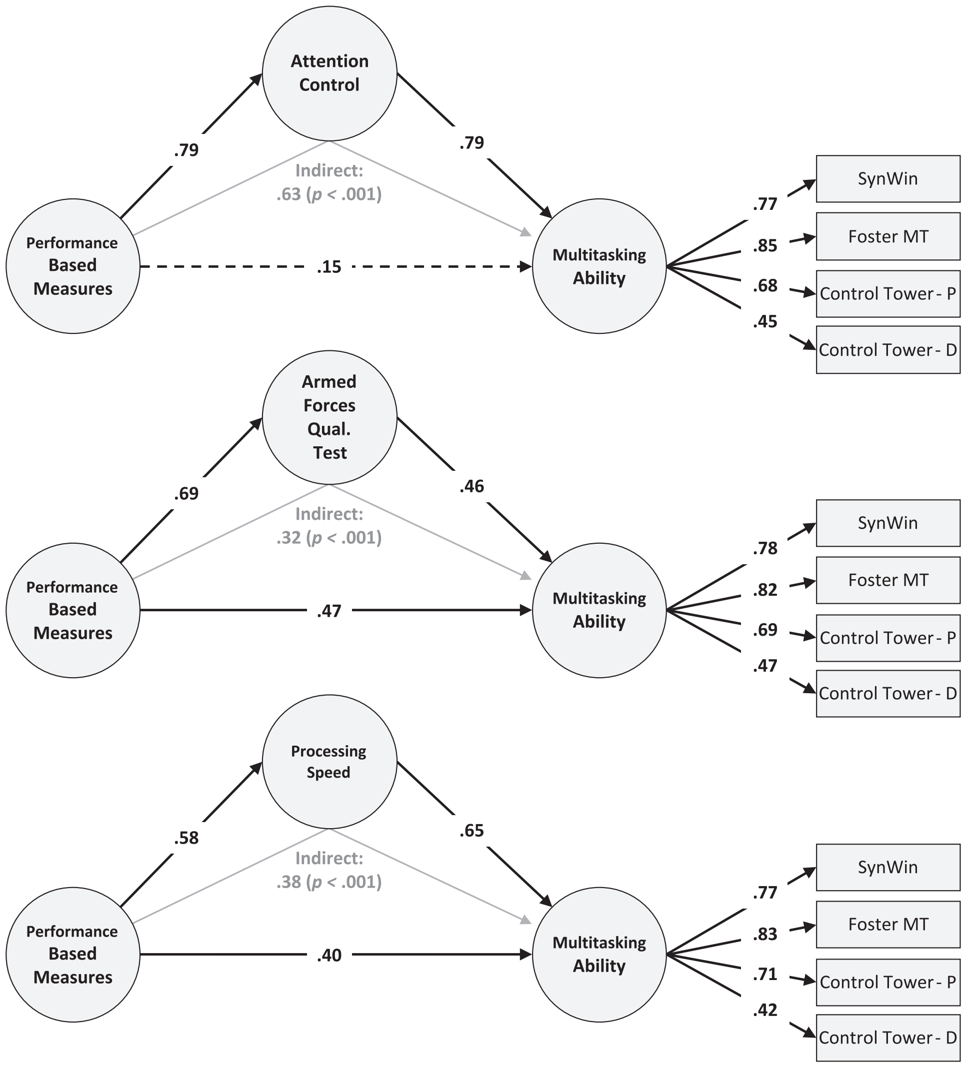

Fig. 12. Mediation models for the effect of performance based measures on multitasking.
Structural equation model with the relationship between the Performance Based Measures and multitasking ability being mediated by attention control (top panel), the Armed Forces Qualification Test (middle panel), and processing speed (bottom panel). Some indicators are not shown for visual clarity.

Table 12
Fit indices for models with attention control mediating the relationship between each performance based measures subtest and multitasking performance.

Performance Based Measures Subtest	χ^2(df), *p*-value	CFI	TLI	RMSEA [90% CI]	SRMR
Terrain Orientation	47.068(18), $p < .001$	0.966	0.946	0.071[0.046,0.096]	0.043
Dichotic Listening Score	43.091(18) $p < .001$	0.971	0.955	0.066[0.041,0.091]	0.033
Two-dimensional Airplane Tracking Score	50.492(18), $p < .001$	0.961	0.940	0.075[0.051,0.100]	0.037
Vertical Tracking Score	43.691(18), $p < .001$	0.969	0.952	0.067[0.042,0.092]	0.035
Emergency Scenarios Score	53.449(18), $p < .001$	0.958	0.934	0.078[0.055,0.103]	0.041

CFI = Confirmatory Fit Index. TLI = Tucker-Lewis Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardized Root Mean Square Residual.

differences. Further, processing speed, which had incremental prediction of simulated work multitasking over attention control and the Armed Forces Qualification Test for predicting multitasking, also had very small gender differences in these data. Building off these findings could lead to more female-identifying applicants being selected, provided these results withstand replication and extension with different samples and criterion constructs (Burgoyne, Mashburn, & Engle, 2021).

3.3. Summary of Study 2

Study 2 addressed validity concerns raised by the online administration of knowledge tests in Study 1 by administering all tests in a proctored, laboratory environment. It also included a broader array of multitasks than Study 1, and it introduced the Performance Based Measures subtest of the ASTB as a covariate to both attention control and the Armed Forces Qualification Test for predicting simulated work performance. Finally, it also controlled for other sources of variation (working memory capacity and processing speed) to better isolate attention control. Contrary to Study 1, the Armed Forces Qualification Test uniquely predicted simulated work multitasking performance ($\beta = 0.36$, $p < .001$) when participants were tested in-lab and after controlling for attention control, processing speed, and the Performance Based

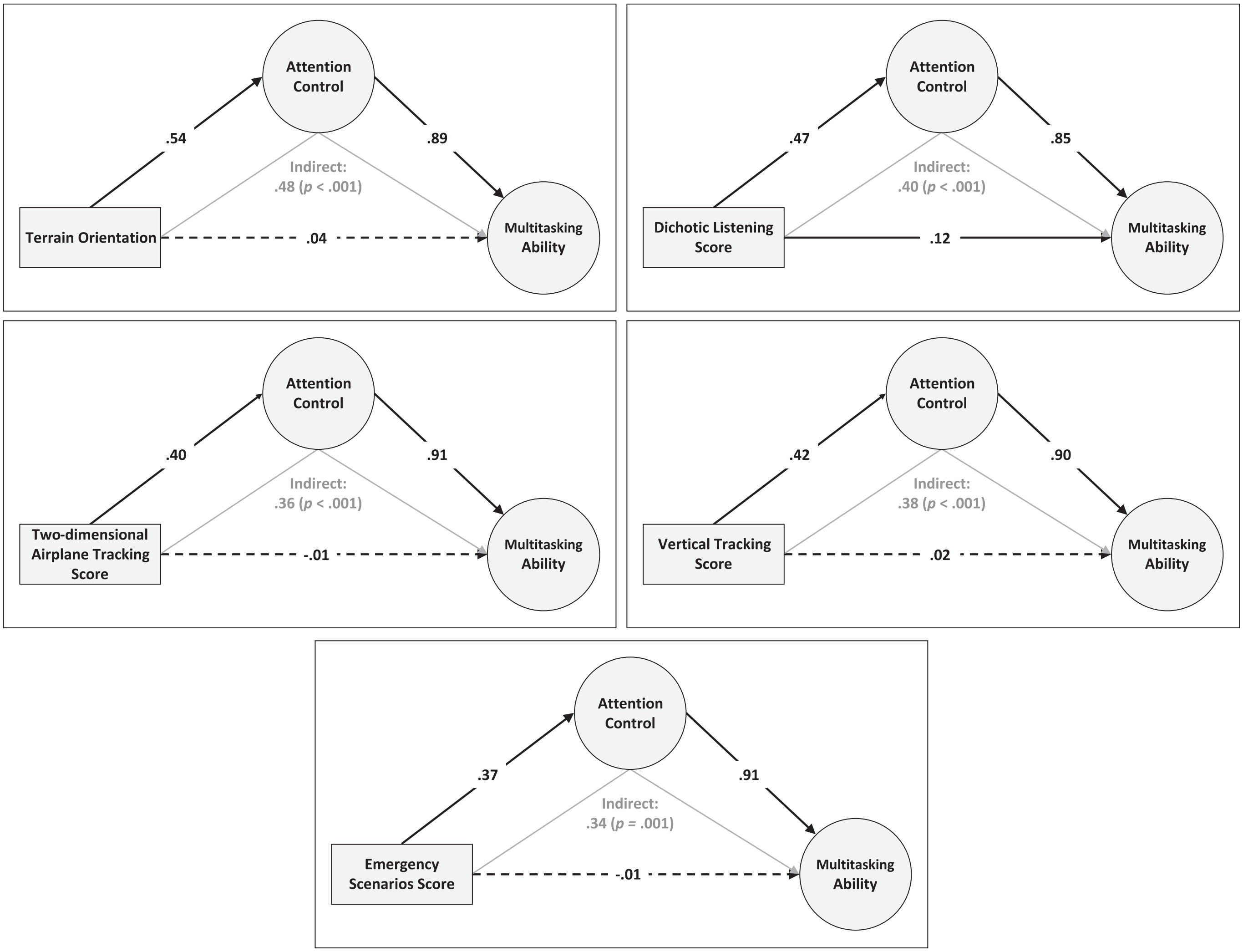

Fig. 13. Attention control mediating the relationship between each of the five performance based measures and multitasking ability.
Dashed lines are not significant at the $p = .05$ level. Factor indicators are omitted for visual clarity. Fit statistics are reported in Table 12.

Measures. Further, while accounting for working memory capacity and processing speed reduced the relationship between attention control and multitasking, it remained a significant and substantial unique predictor across all analyses. The Performance Based Measures, however, did not. Attention control, but not the Armed Forces Qualification Test or processing speed, fully mediated the relationship between the Performance Based Measures and multitasking performance; this was true both at the latent level and for four of the five subtasks. This suggests that the Performance Based Measures might tap attention control processes by proxy rather than merely reflecting psychomotor skill. Finally, we also observed relatively small gender differences in performance on the attention control tasks and the processing speed tasks, suggesting that their use could lead to the more equitable selection of non-male-identifying applicants in applied contexts.

4. General discussion

Across two studies (total $N \approx 800$), we investigated the relationships between attention control, acquired knowledge, psychomotor ability, and performance in simulated work multitasking paradigms. Study 1 investigated the predictive validity of attention control and two primarily knowledge-based tests, the Wonderlic and the Armed Forces Qualification Test. At the latent level, attention control accounted for a majority of the reliable variance in simulated work performance on its own ($R^2 = 55.0\%$). Adding the Wonderlic and the Armed Forces Qualification Test did not increase the total R^2 (i.e., $R^2 = 54.7\%$), and in the full structural equation model, only the contribution of attention control to simulated work multitasking performance was statistically significant ($\beta = 0.70$, $p < .001$).

These results are somewhat surprising because the SimWork multitasking paradigm included a mathematics component, and both the Armed Forces Qualification Test and the Wonderlic have a substantial number of mathematics items. Although we anticipated this would drive some relationship between the math-based predictors and multitasking performance, ultimately attention control dominated, eclipsing the contribution of the Armed Forces Qualification Test and the Wonderlic. One interpretation of this result is that SimWork primarily reflects individual differences in the ability to maintain focus and avoid distraction and interference while managing multiple ongoing tasks. Even though the multitask demanded domain-specific math knowledge, domain-general attention control was the primary source of individual differences. As a counterpoint, it should be noted that the Armed Forces Qualification Test includes also tests of verbal ability, possibly reducing its predictive validity for the simulated work measure used in this study. Indeed, the two math subtests of the Armed Forces Qualification Test had numerically larger correlations with overall simulated work performance ($rs = 0.36$ and 0.30) than the two verbal subtests ($r = 0.25$ and 0.23). This explanation of our results is unlikely, however, since

Table 13
Subgroup differences in performance by gender.

Measure	Men			Women			Difference
	n	*Mean*	*SD*	*n*	*Mean*	SD	Cohen's *d* [95% CI]
SynWin	119	3395.44	510.09	175	3169.03	540.04	**−0.43 [−0.67, −0.20]**
FosterMT	116	101,853.35	23,855.87	174	92,442.82	27,468.69	**−0.37 [−0.60, −0.13]**
Control Tower (P)	123	104.08	29.40	183	99.33	32.76	−0.15 [−0.38, 0.08]
Control Tower (D)	119	26.21	2.31	177	25.57	2.58	**−0.26 [−0.50, −0.03]**
Flanker Squared	116	43.83	14.24	173	37.48	13.49	−0.46 [−0.70, 0.22]
Stroop Squared	124	42.18	12.79	185	40.51	12.63	−0.13 [−0.36, 0.10]
Simon Squared	124	70.46	9.25	184	65.98	9.14	**−0.49 [−0.72, −0.26]**
AFQT Composite	108	94.84	11.61	170	89.61	10.65	**−0.47 [−0.72, −0.22]**
AFQT Arithmetic Reasoning	108	18.13	3.15	170	15.37	4.09	**−0.76 [−1.01, −0.51]**
AFQT Math Knowledge	108	15.94	2.05	170	14.27	2.77	**−0.69 [−0.93, −0.44]**
AFQT Word Knowledge	108	20.25	3.22	170	19.95	2.89	−0.10 [−0.34, 0.14]
AFQT Paragraph Comprehension	108	10.14	1.58	170	10.04	1.49	−0.06 [−0.31, 0.18]
PBM Terrain Orientation	47	0.52	0.24	87	0.40	0.23	**−0.51 [−0.88, −0.15]**
PBM Dichotic Listening Score	97	25.33	11.29	152	24.32	10.05	−0.09 [−0.35, 0.16]
PBM Two-dimensional Airplane Tracking Score	94	11.89	8.77	151	4.40	4.83	**−1.06 [−1.35, −0.77]**
PBM Vertical Tracking Score	97	28.98	9.30	150	22.41	9.50	**−0.70 [−0.96, −0.43]**
PBM Emergencies Score	97	10.84	10.95	152	9.10	9.43	−0.17 [−0.43, 0.09]
Digit Comparison	121	30.46	5.70	175	29.43	5.30	−0.19 [−0.42, 0.05]
Letter Comparison	121	20.73	4.10	175	20.39	4.12	−0.08 [−0.31, 0.15]
Pattern Comparison	120	39.32	6.05	175	38.71	5.90	−0.10 [−0.33, 0.13]

n = sample size, *SD* = standard deviation. Negative Cohen's *d* values indicate that the male subgroup scored higher on the measure than the female subgroup. Bolded estimates of Cohen's *d* indicate that the group difference was statistically significant at the $\alpha < 0.05$ level, according to a Welch's *t*-test. Control Tower (P) = Primary, Control Tower (D) = Distractor. AFQT = Armed Forces Qualification Test. PBM = Performance Based Measures subtest of the Aviation Selection Test Battery.

mathematics subtests had larger factor loadings on the Armed Forces Qualification Test latent variable than the verbal subtests. Another reason for the knowledge tests' null predictive relationship might be that the online administration of the Armed Forces Qualification Test and Wonderlic allowed participants to look up answers on the tests, undermining their validity.

Study 2 focused on attention control, the Armed Forces Qualification Test, and the Performance Based Measures, a subtest of the Aviation Selection Test Battery involving manual and bimanual tracking as well as dichotic listening, mental rotation, and following instructions. It was conducted in-lab to address concerns about cheating on the acquired knowledge tests from Study 1. It also statistically controlled for working memory capacity and processing speed to isolate variance specific to attention control. Confirmatory factor analysis revealed strong correlations between multitasking ability and attention control ($r = 0.91$), the Performance Based Measures ($r = 0.79$), the Armed Forces Qualification Test ($r = 0.62$), and processing speed ($r = 0.89$). The correlation between working memory capacity and processing speed was smaller at $r = 0.52$. Controlling for processing speed reduced but did not eliminate the relationship between attention control and multitasking, but the relationship between working memory capacity and multitasking was no longer significant once attention control and processing speed were controlled for. In the full structural equation model, only attention control ($\beta = 0.31$, $p < .05$), the Armed Forces Qualification Test ($\beta = 0.36$, $p < .001$), and processing speed ($\beta = 0.45$, $p < .001$) significantly predicted multitasking ability, whereas the Performance Based Measures ($\beta = 0.02$, $p = .86$) did not.

The significant unique relationship between the Armed Forces Qualification Test and multitasking in Study 2 may support our concerns about the validity of the online knowledge tests used in Study 1. This is because the test was predictive in a proctored testing environment, but not when taken online in an un-proctored testing environment, even after controlling for attention control and the Wonderlic. However, scores for all tasks that were shared across both studies were numerically larger in Study 2 than in Study 1 (compare Tables 3 and 8). This could indicate that the Study 2 sample may be somewhat higher in general cognitive ability than Study 1, which may explain why the Armed Forces Qualification Test was predictive in Study 2 but not Study 1. It is possible that the Armed Forces Qualification Test items might provide greater test information for moderate to high ability samples, a possibility that could be explored in future work by fitting a polytomous Item Response Theory model to the present data.

However, even if this were the case, attention control was a strong predictor of multitasking across both samples. This point warrants elaboration. Given the large correlations among our latent predictors in Study 2 (see Table 10), they likely account for overlapping variance in simulated work performance. Indeed, the unique predictive paths controlling for other variables are much smaller than the uncontrolled between-factor correlations (compare Table 10 and Fig. 11). The relationship between attention control and multitasking was robust to a number of statistical controls despite having substantial overlap with other predictors. However, results from Study 2 suggests that the strong relationship between attention control and multitasking observed in Study 1 would likely be smaller if other constructs, such as processing speed, were considered.

Turning to the Performance Based Measures, mediation analyses revealed that the relationship between the Performance Based Measures and multitasking ability was fully explained by attention control, but not the Armed Forces Qualification Test or processing speed. Even at the subtest-level, attention control fully mediated the relationship between four of the five measures from the Performance Based Measures and multitasking ability and partially mediated the fifth. Thus, attention control largely accounted for the validity of the Performance Based Measures for predicting simulated work multitasking performance.

This result was unexpected because the Performance Based Measures and the multitasking paradigms used in the present study both require psychomotor ability. For example, in the Performance Based Measures, participants must coordinate the use of hands-on-throttle-and-stick controllers for most subtests. In the multitasking paradigms, participants must quickly and accurately direct mouse movements according to rapidly changing task demands. Even so, nearly all the predictive validity of the Performance Based Measures was accounted for by attention control. One reason for this finding might be that the psychomotor demands in our simulated work multitasks are comparatively minor to that of the Performance Based Measures and to that of actual Naval training, such as pilot primary school (see Burgoyne, Mashburn, et al., n.d., under review). Psychomotor ability did not predict simulated work multitasking performance beyond attention control, but these results may differ given a different criterion measure.

Evaluating the cognitive demands imposed by the Performance

Based Measures helps make sense of the full mediation observed here. For instance, the tracking subtasks require attending to moving targets and crosshairs while monitoring one's motor behaviors to reduce the distance between the targets and crosshairs. Although there is clearly a motoric component to the task(s) (e.g., moving the throttle or stick), this may not explain the relationship between the Performance Based Measures and at least some complex task performance. Attending to stimuli, resisting internal and external distractions, and monitoring the consequences of one's actions are all theoretically supported by attention control (Burgoyne & Engle, 2020). Furthermore, as most of the Performance Based Measures require multitasking (e.g., performing Two-dimensional Airplane Tracking and Vertical Tracking simultaneously, while engaged in a dichotic listening task), this likely increases the attentional demand of the tasks. Thus, the primary "active ingredient" common to both the Performance Based Measures and the simulated work multitasks used in the present study may be the ability to control one's attention in a goal-directed fashion.

If true, there may be several practical advantages to using tests of attention control rather than the Performance Based Measures for predicting complex task performance, particularly if these results were to be replicated in Navy samples when predicting actual training and/or job performance. For instance, the Performance Based Measures currently requires the use of hands-on-throttle-and-stick controllers (see Fig. 6), a set of peripheral devices with which many test-takers will be inexperienced. Test-takers with more experience using a hands-on-throttle-and-stick controllers (e.g., video game players) are likely to have a significant advantage over those lacking experience, potentially confounding the validity of the measure for individual differences research (Sibley, Herdener, Coyne, Drollinger, & Strong, 2023). By contrast, attention control measures can be administered via standard computers or tablets without extra hardware (Burgoyne, Tsukahara, Mashburn, Pak, & Engle, 2023). This would also reduce testing costs and increase testing flexibility. Additionally, the Performance Based Measures requires at least 30 min of testing time, whereas our three-minute tests of attention control (i.e., the "Squared" tasks) are much more efficient. Finally, the Performance Based Measures may have larger gender differences than tests of attention control, although this finding should be replicated in larger samples. If these general patterns hold in other similar investigations, it might suggest that including attention control tests in selection contexts may increase the number of female-identifying applicants selected for certain vocations and training programs.

At a more theoretical level, this investigation is consistent with other, similar studies of multitasking. For example, Redick et al. (2016) can be seen as a precursor to the present study; it established that several of the multitasks we used in Study 2 load on a common factor with psychometrically-sound properties. Moreover, it included as predictors of multitasking ability multiple measures of working memory, attention control, short-term storage, as well as fluid intelligence. Directly relevant to our results, Redick et al. (2016) found that measures of attention control and measures of short-term storage capacity together fully mediated the relation between working memory capacity and multitasking performance. Where our results extend Redick et al.'s (2016) is through the use of new-and-improved measures of attention control (i.e., the three "Squared" tests; Burgoyne, Tsukahara, Mashburn, Pak, & Engle, 2023) and the inclusion of processing speed measures. Many of the attention control measures used by Redick et al. (2016) had relatively low loadings on their common factor (i.e., range: 0.25 to 0.65, with an average loading of 0.46). For comparison, the three attention control tasks used in the present studies had loadings ranging from 0.64 to 0.80 across two studies. Thus, the attention control factor used here was able to capture more of the reliable variance among the observed measures than is often observed in studies of attentional abilities (see, e.g., Draheim, Mashburn, Martin, & Engle, 2019). Additionally, we were able to isolate attention control from processing speed despite a strong correlation between them, providing further clarity on the unique contributions of these constructs to multitasking performance.

Finally, readers may wonder about our operationalization of attention control in the current studies, especially as it pertains to other concepts in the executive functioning literature. In particular, the Stroop Squared, Flanker Squared, and Simon Squared tasks are all modifications of tasks normally thought to capture the "inhibition" subfactor of executive functioning (Miyake et al., 2000). If we are using inhibition measures, why insist on using the term "attention control?" In short, we do not think that the two ideas are neatly separable, but rather prefer to think of "inhibition" as one way attention may be directed. The present attention control tasks reflect this in their design. In addition to stimulus-stimulus and stimulus-response conflict, which are characteristic of inhibition measures, our attention control tasks have a switching element whereby the stimulus dimension that was irrelevant when evaluating the target stimulus becomes relevant when making a response. This added switching demand means that our attention control tasks more likely reflect the "common factor" of executive functioning than inhibition, specifically (c.f., Friedman & Miyake, 2017). Thus, we favor the term "attention control."

4.1. Strengths and limitations

The present investigation has several strengths. A major strength is the relatively large, diverse sample of both the online and in-lab studies. Also, we included multiple measures of each construct whenever possible. Combined, these design features increase the generalizability of our main findings. However, there are also weaknesses to our approach.

As we have already suggested, one limitation of this work is our use of simulated work multitasking paradigms as a proxy for real-world work. Despite multitasking's relevance to real-world contexts (Barron & Rose, 2017), real-world jobs likely impose additional cognitive demands beyond those of our simulated work tasks. For instance, acquired knowledge is critical for some jobs (e.g., practicing law) and should be integral to selecting individuals for such roles. In those cases, knowledge measures would likely demonstrate greater validity than was observed in the present studies, particularly to the extent that knowledge predictors are matched in bandwidth and fidelity to criterion measures (Brunswik, 1952). Additionally, jobs with more complex psychomotor demands might show a different relationship with the Performance Based Measures than the one observed with our simulated work multitasks. For example, the degree of psychomotor control that is demanded when piloting an aircraft likely greatly exceeds the psychomotor demands of our point-and-click multitasks.

That said, many jobs require fluid abilities in addition to domain-specific knowledge. For example, piloting an aircraft requires knowledge of the cockpit's control panel, however, it also requires pilots to attend to a considerable amount of incoming information, monitor their performance, and respond to emergency scenarios in real-time. In these situations, we would expect both knowledge and more fluid abilities, such as the ability to control attention, to be important predictors of performance. Indeed, in other work, we have shown that attention control predicts aviation training outcomes beyond existing recruitment tests, including the Armed Forces Qualification Test and Aviation Selection Test Battery, of which the Performance Based Measures is a part (Burgoyne, Mashburn, et al., n.d., under review). In future work, we hope to continue this research using other samples of military personnel to investigate how predictors of performance differ as a function of occupational demands.

4.2. Conclusion and general audience summary

This work explored the relative contributions of knowledge, psychomotor ability, and attention control to performance in complex simulated work multitasking paradigms. We found that attention control was a major predictor of performance above and beyond acquired knowledge measures that are used to select personnel in the occupational sector (i.e., the Wonderlic and the Armed Forces Qualification Test). We also found that attention control fully mediated the relationship between a test of psychomotor abilities and spatial skill used by the U.S. Military, the Aviation Selection Test Battery's Performance Based Measures, and simulated work multitask performance. Future work should investigate other psychomotor measures to see if attentional abilities also underpin their predictive validity. Additionally, future research should extend this investigation into military samples with real-world training and performance data to determine whether attention control also mediates the relationship between the Performance Based Measures and training outcomes. If so, the attention control tasks presented here may be a resource-efficient alternative or supplement to some standard ability tests in military and occupational contexts.

Author note

A pre-print of this manuscript was uploaded to PsyArXiv (https://psyarxiv.com/gdmtf) on June 6th, 2023. Data are not currently approved for public release. We thank Sarah Melick for her guidance on the Aviation Selection Test Battery. Questions about this manuscript should be directed to Cody A. Mashburn (cmashburn3@gatech.edu).

Funding

This work was supported by a Visionary Grant from the American Psychological Foundation to Alexander P. Burgoyne, Naval Research Laboratory Grants N00173–20-2-C003 and N00173–20-P-0135 to Randall W. Engle, by award P14278 from the Institute of Mental Chronometry to the Georgia Tech Research Corporation.

CRediT authorship contribution statement

Cody A. Mashburn: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Alexander P. Burgoyne:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jason S. Tsukahara:** Software, Methodology, Data curation. **Richard Pak:** Software, Methodology, Data curation. **Joseph T. Coyne:** Writing – review & editing, Supervision, Software, Resources, Methodology, Funding acquisition, Data curation. **Ciara Sibley:** Writing – review & editing, Supervision, Software, Resources, Methodology, Investigation, Conceptualization. **Cyrus Foroughi:** Writing – review & editing, Project administration, Investigation, Funding acquisition. **Randall W. Engle:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Alexander P. Burgoyne reports financial support was provided by American Psychological Foundation. Randall W. Engle reports financial support was provided by US Naval Research Laboratory.

Data availability

The authors do not have permission to share data.

Appendix A. Appendix

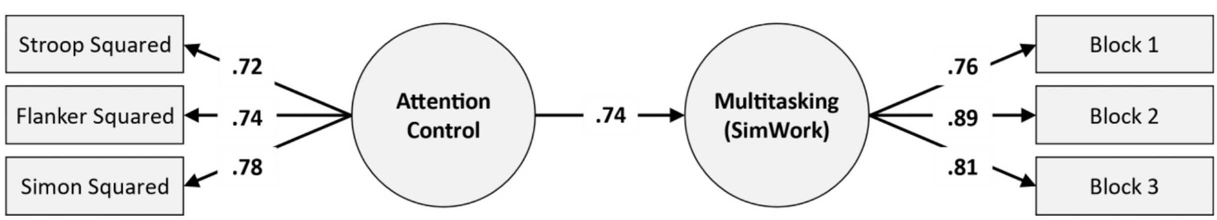

Fig. A1. Attention control predicting simulated work performance from Study 1
Structural equation model with the relationship between attention control and simulated work multitasking ability. Model fit was excellent, χ^2 (8) = 17.01, p = .030, CFI = 0.990, TLI = 0.982, RMSEA = 0.050, 90% CI [0.015 0.084], SRMR = 0.022.

Appendix B. Appendix

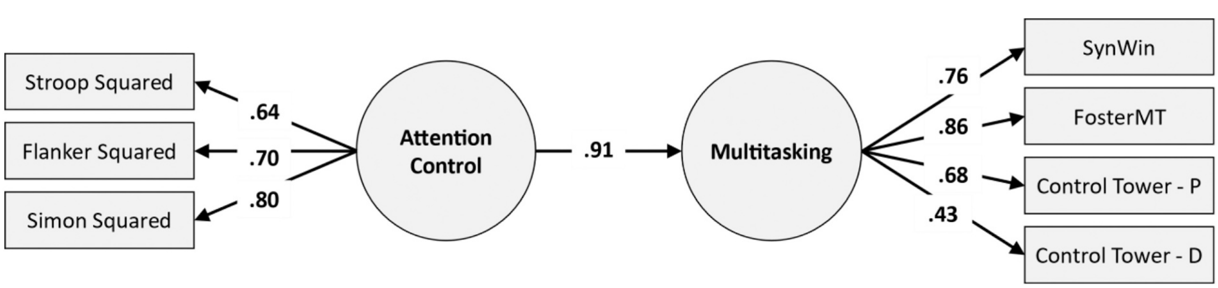

Fig. B1. Attention control predicting simulated work performance from Study 2.
Structural equation model with the relationship between attention control and simulated work multitasking ability. Model fit was acceptable, χ^2 (13) = 37.548, p < .001; CFI = 0.969, TLI = 0.951, RMSEA = 0.077, 90% CI [0.049, 0.106], SRMR = 0.033.

References

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.

Allen, R. J., Waterman, A. H., Yang, T. X., & Jaroslawska, A. J. (2022). *Working memory in action: Remembering and following instructions. Memory in science for society: There is nothing as practical as a good theory*.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.

Barron, L. G., & Rose, M. R. (2017). Multitasking as a predictor of pilot performance: Validity beyond serial single-task assessments. *Military Psychology, 29*(4), 316–326.

Baumeister, R. F., Schmeichel, B. J., & Vohs, K. D. (2007). Self-regulation and the executive function: The self as controlling agent. *Social Psychology: Handbook of Basic Principles, 2*, 516–539.

Berry, C. M., Gruys, M. L., & Sackett, P. R. (2006). Educational attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology, 91*(3), 696.

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.

Bosco, F., Allen, D. G., & Singh, K. (2015). Executive attention: An alternative perspective on general mental ability, performance, and subgroup differences. *Personnel Psychology, 68*(4), 859–898.

Brunswik, E. (1952). *The conceptual framework of psychology. (Int. Encycl. unified Sci., v. 1, no. 10.)*.

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

Burgoyne, A. P., Mashburn, C. A., & Engle, R. W. (2021). Reducing adverse impact in high-stakes testing. *Intelligence, 87*, Article 101561.

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.

Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., Pak, R., Coyne, J. T., Foroughi, C., Sibley, C., Drollinger, S. M., & Engle, R. W. (n.d.) (under review). Attention control measures improve the prediction of performance in Navy trainees.

Burgoyne, A. P., Seeburger, D. T., & Engle, R. W. (n.d.) (in press). Modality matters: Three auditory conflict tasks to measure individual differences in attention control. Behavior Research Methods.

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). Nature and measurement of attention control. *Journal of Experimental Psychology: General, 152*(8), 2369–2402.

Buszard, T., Farrow, D., Verswijveren, S. J., Reid, M., Williams, J., Polman, R., … Masters, R. S. (2017). Working memory capacity limits motor learning when implementing multiple instructions. *Frontiers in Psychology, 8*, 1350.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1.

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. Elsevier Science Publishers B.V.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25*(5), 975–979.

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

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163.

Cowan, N., Morey, C. C., & Naveh-Benjamin, M. (2020). An embedded-processes approach to working memory: How is it distinct from other approaches, and to what ends? In R. H. Logie, V. Camos, & N. Cowan (Eds.), *Working Memory: The State of the Science* (pp. 44–84).

Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What Item Response Theory can tell us about the complex span tasks. *Psychological Assessment, 30*(1), 116–129.

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*, 508–535.

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*, 1143–1197.

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.

Elsmore, T. F. (1994). SYNWORK: APC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instrumentation, and Computers, 26*, 421–426.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23.

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

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *The Journal of Educational Research, 84*(5), 253–262.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128*(3), 309.

Fatolitis, P. G., Jentsch, F. G., Hancock, P. A., Kennedy, R. S., & Bowers, C. (2010). *Initial validation of novel performance-based measures: Mental rotation and psychomotor ability*. Naval Aerospace Medical Research Laboratory.. https://apps.dtic.mil/sti/citations/ADA529481

Fleishman, E. A. (1972). On the relation between abilities, learning, and human performance. *American Psychologist, 27*(11), 1017.

Fleishman, E. A., & Rich, S. (1963). Role of kinesthetic and spatial-visual abilities in perceptual-motor learning. *Journal of Experimental Psychology, 66*(1), 6.

Fleishman, E. A., Teichner, W. H., & Stephenson, R. W. (1970). *Development of a taxonomy of human performance: A review of the second year's progress*.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Furnham, A. (2012). Learning style, personality traits and intelligence as predictors of college academic performance. *Individual Differences Research, 10*(3).

Gibb, G. D., & Dolgin, D. L. (1988, October). Validation of a computer-based aviation secondary selection system for student naval aviators. In *Proceedings of the human factors society annual meeting (Vol. 32, no. 14, pp. 807-811)*. Los Angeles, CA: SAGE Publications.

Gonthier, C. (2023). An easy way to improve scoring of memory span tasks: The edit distance, beyond "correct recall in the correct serial position". *Behavior Research Methods, 55*(4), 2021–2036.

Hambrick, D. Z., Burgoyne, A. P., Altmann, E. M., & Matteson, T. J. (2023). Explaining the validity of the ASVAB for job-relevant multitasking performance: The role of placekeeping ability. *Journal of Intelligence, 11*(12), 225.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology, 24*(8), 1149–1167.

Hambrick, D. Z., Rench, T. A., Poposki, E. M., Darowski, E. S., Roland, D., Bearden, R. M., … Brou, R. (2011). The relationship between the ASVAB and multitasking in navy sailors: A process-specific approach. *Military Psychology, 23*(4), 365–380.

Hicks, K. L., Harrison, T. L., & Engle, R. W. (2015). Wonderlic, working memory capacity, and fluid intelligence. *Intelligence, 50*, 186–195.

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.

Hunter, J. E. (2017). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In *Performance measurement and theory* (pp. 257–266). Routledge.

Jaroslawska, A. J., Gathercole, S. E., Logie, M. R., & Holmes, J. (2016). Following instructions in a virtual school: Does working memory play a role? *Memory & Cognition, 44*, 580–589.

JASP Team. (2023). *JASP (Version 0.17.1) [Computer software]*.

Kvist, A. V., & Gustafsson, J. E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's investment theory. *Intelligence, 36*(5), 422–436.

LePine, J. A., Colquitt, J. A., & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel Psychology, 53*(3), 563–593.

Logan, G. D., & Crump, M. J. (2011). Hierarchical control of cognitive processes: The case for skilled typing. In B. H. Ross (Ed.), *The psychology of learning and motivation (Vol. 54)* (pp. 1–27).

Martin, J., 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.

Mashburn, C. A., Burgoyne, A. P., & Engle, R. W. (2023). Working memory, intelligence, and life success: Examining relations to academic achievement, job performance, physical health, mortality, and psychological well-being. In R. H. Logie, N. Cowan, S. E. Gathercole, R. W. Engle, & Z. E. Wen (Eds.), *Memory in science for society: There is nothing as practical as a good theory*.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Meier, M. E., Smeekens, B. A., Silvia, P. J., Kwapil, T. R., & Kane, M. J. (2018). Working memory capacity and the antisaccade task: A microanalytic–macroanalytic investigation of individual differences in goal activation and maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(1), 68–84.

Melton, A. W. (1947). *Apparatus tests (Army air forces aviation psychology program research report No. 4)*. United States Government Printing Office.

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/co gp.1999.0734

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*(1), 56–60.

Neisser, U. (1964). Visual search. *Scientific American, 210*(6), 94–102.

Nye, C. D., Chernyshenko, O. S., Stark, S., Drasgow, F., Phillips, H. L., Phillips, J. B., & Campbell, J. S. (2020). More than g: Evidence for the incremental validity of performance-based assessments for predicting training performance. *Applied Psychology, 69*(2), 302–324.

Ohi, K., Takai, K., Sugiyama, S., Kitagawa, H., Katoaka, Y., Soda, M., Kitaich, K., Kawasaki, Y., Ito, M., & Shiori, T. (2022). Intelligence declines across major depressive disorder, bipolar disorder, and schizophrenia. *CNS Spectrums, 27*(4), 468–474.

Powers, R. (2011). *ASVAB for dummies* ((3rd ed.).). Wiley Publishing, Inc.

Rakhmanov, O., & Dane, S. (2021). Improving the prediction accuracy of academic performance of the freshman using Wonderlic personnel test and Rey-Osterrieth complex figure. In *Information and communication technology and applications: Third international conference, ICTA 2020, Minna, Nigeria, November 24–27, 2020, revised selected papers 3 (pp. 54-65)*. Springer International Publishing.

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: Selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37* (2), 308–324.

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

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.

Roberts, R. D., Goff, G. N., Anjoul, F., Kyllonen, P. C., Pallier, G., & Stankov, L. (2000). The armed services vocational aptitude battery (ASVAB): Little more than acculturated learning (Gc)!? *Learning and Individual Differences, 12*, 81–103.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*(5), 763–776.

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.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*(1), 162.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*(3), 432.

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology, 73*(1), 46.

Schweizer, K., & Koch, W. (2002). A revision of Cattell's investment theory: Cognitive properties influencing learning. *Learning and Individual Differences, 13*(1), 57–82.

Sibley, C., Herdener, N., Coyne, J., Drollinger, S., & Strong, K. (2023, June 2). Exploring practice effects in the Navy's aviation selection test [poster presentation]. In *22nd international symposium on aviation psychology. Rochester, New York, United States*.

Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition, 4*, 215–230.

Trent, J. D., & Aguilar, I. D. (2020). *Validation of the pilot candidate selection method (PCSM) (air force personnel center strategic research and assessment branch report no. AFCAPS-TR-2020-0002)*. Defense Technical Information Center. Retrieved from https://apps.dtic.mil/sti/citations/AD1101783.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104–132.

Unsworth, N., Robison, M. K., & Miller, A. L. (2021). Individual differences in lapses of attention: A latent variable analysis. *Journal of Experimental Psychology: General, 150* (7), 1303–1331.

Walker, P. B., Olde, B. A., & Olson, T. M. (2007, January). Improving aviator selection using the performance-based measurement battery (PBMB). Poster presented at the proceedings of the human systems integration symposium, Annapolis, Maryland. Retrieved from https://www.researchgate.net/publication/266910869_Improving_Aviator_Selection_Using_the_Performance-Based_Measurement_Battery_PBMB.

Zelazo, P. D., & Cunningham, W. A. (2007). *Executive function: Mechanisms underlying emotion regulation*.