


Article

Explaining the Validity of the ASVAB for Job-Relevant Multitasking Performance: The Role of Placekeeping Ability

David Z. Hambrick^{1,*}, Alexander P. Burgoyne², Erik M. Altmann¹ and Tyler J. Matteson³¹ Department of Psychology, Michigan State University, East Lansing, MI 48824, USA² School of Psychology, Georgia Institute of Technology, Atlanta, GA 30327, USA³ Department of Psychology, Stanford University, Stanford, CA 94305, USA

* Correspondence: hambric3@gmail.com

Abstract: Scores on the Armed Services Vocational Aptitude Battery (ASVAB) predict military job (and training) performance better than any single variable so far identified. However, it remains unclear what factors explain this predictive relationship. Here, we investigated the contributions of *fluid intelligence* (Gf) and two executive functions—*placekeeping ability* and *attention control*—to the relationship between the Armed Forces Qualification Test (AFQT) score from the ASVAB and job-relevant multitasking performance. Psychometric network analyses revealed that Gf and placekeeping ability independently contributed to and largely explained the AFQT–multitasking performance relationship. The contribution of attention control to this relationship was negligible. However, attention control did relate positively and significantly to Gf and placekeeping ability, consistent with the hypothesis that it is a cognitive “primitive” underlying the individual differences in higher-level cognition. Finally, hierarchical regression analyses revealed stronger evidence for the incremental validity of Gf and placekeeping ability in the prediction of multitasking performance than for the incremental validity of attention control. The results shed light on factors that may underlie the predictive validity of global measures of cognitive ability and suggest how the ASVAB might be augmented to improve its predictive validity.

Keywords: ASVAB; cognitive ability; job performance; individual differences; placekeeping ability; attention control



Citation: Hambrick, David Z., Alexander P. Burgoyne, Erik M. Altmann, and Tyler J. Matteson. 2023. Explaining the Validity of the ASVAB for Job-Relevant Multitasking Performance: The Role of Placekeeping Ability. *Journal of Intelligence* 11: 225. <https://doi.org/10.3390/jintelligence11120225>

Received: 1 July 2023

Revised: 16 November 2023

Accepted: 2 December 2023

Published: 15 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Decades of research have established that scores on standardized tests of cognitive ability predict job (and training) performance to a statistically and practically significant degree (Schmidt and Hunter 2004). One such test is the Armed Services Vocational Aptitude Battery (ASVAB). Taken annually by an estimated one million U.S. high school students, the ASVAB is the U.S. military’s primary selection and placement instrument. The ASVAB includes nine subtests (see Table 1). Criterion correlations average around 0.60 after correction for measurement error and range restriction (e.g., Mayberry 1990; Welsh et al. 1990; Ree and Earles 1992; Ree et al. 1994; Larson and Wolfe 1995; Held and Wolfe 1997) and are stable across different levels of job experience (Schmidt et al. 1988; Reeve and Bonaccio 2011; Hambrick et al. 2023). Economic utility analyses indicate that using the ASVAB for military selection saves the U.S. military (and taxpayers) billions of dollars per year (Schmidt et al. 1979; Hunter and Schmidt 1996; Held et al. 2014).

Nevertheless, it remains unclear *why* ASVAB scores predict job performance. That is, what aspects of the cognitive system does the ASVAB measure that explain its predictive validity? Presumably, part of the answer to this question is that the ASVAB captures skills and knowledge that are directly transferable to tasks in the military workplace. For example, the Paragraph Comprehension subtest measures skill in comprehending written text, which is important for work tasks such as learning how to perform a procedure from an instructional manual. As another example, the Mechanical Knowledge subtest measures

knowledge of physical and mechanical principles, which is directly relevant to tasks that a mechanic or engineer performs.

Table 1. Description of ASVAB Subtests.

Subtest	Acronym	Description	Items	Time
General Science	GS	Knowledge of physical and biological sciences	25	11
Arithmetic Reasoning *	AR	Ability to solve arithmetic word problems	30	36
Word Knowledge *	WK	Knowledge of and ability to infer word meanings	35	11
Paragraph Comprehension *	PC	Ability to obtain information from written passages	15	13
Mathematics Knowledge *	MK	Knowledge of high school-level mathematics concepts	25	24
Electronics Information	EI	Knowledge of electrical principles and electronic devices	20	9
Automotive and Shop Information	AS	Knowledge of automotive technology, shop terminology	25	11
Mechanical Comprehension	MC	Knowledge of mechanical and physical principles	25	19
Assembling Objects	AO	Ability to imagine how objects look when disassembled	25	15

Note. *, subtest used to compute the Armed Forces Qualification Test score, using the formula $2(WK + PC) + MK + AR$. Subtest descriptions paraphrased from Wall (2017).

In fact, evidence indicates that the ASVAB primarily measures the broad cognitive ability Cattell (1943) termed *crystallized intelligence* (Gc)—skills and knowledge acquired through experience. For example, in a confirmatory factor analysis, Roberts et al. (2000) found that five of the nine subtests on the current ASVAB loaded with Gc tests from another test battery. However, evidence also indicates that the ASVAB measures other cognitive abilities to some degree, including *fluid intelligence* (Gf)—the ability to solve novel problems. For example, Kyllonen and Christal (1990) and Kyllonen (1993) found that scores on the Arithmetic Reasoning and Mathematics Knowledge ASVAB subtests correlated very highly with a latent factor comprising both reasoning and working memory measures. Similarly, in the above-mentioned study, Roberts et al. found that the Arithmetic Reasoning and Mathematics Knowledge subtests loaded with Gf tests from the other test battery.

There are several possible explanations for the positive relationship between ASVAB scores and Gf. One is that subtests such as Arithmetic Reasoning, Mechanical Comprehension, and Assembling Objects call upon reasoning processes, along with other cognitive abilities and knowledge/skill. Another is that ASVAB scores index variance in Gf because Gf influenced the degree to which people earlier in life, through formal and informal learning experiences, acquired the skills and knowledge the ASVAB assesses (Cattell [1971] 1987).

The ASVAB may also capture *executive functions*: domain-general cognitive abilities involved in effortfully guiding behavior toward goals, especially under non-routine circumstances (Banich 2009). In recent years, two executive functions have been proposed that seem particularly relevant to explaining the predictive validity of the ASVAB. The first is *attention control*: the ability to maintain task-relevant information in a highly active state within working memory, especially under conditions of distraction or interference (Hasher et al. 2007; Burgoyne and Engle 2020). The second is what we have termed *placekeeping ability*: the ability to perform a sequence of mental operations in a prescribed order without skipping or repeating steps (Altmann et al. 2014). Placekeeping performance presumably involves attention control, but may, in addition, capture operations involved in a broad class of tasks with a procedural component in which keeping track of one's place in a set of steps is important.

The ASVAB may measure attention control because, to answer a test item correctly, the test taker presumably must attend to it, and placekeeping ability, as some of the test items require the test taker to follow a procedure to solve a problem (e.g., multiplying binomials). Alternatively, just as they may for Gf, ASVAB scores may index attention control and placekeeping ability, if these influenced the extent to which participants acquired the skills

and knowledge the ASVAB measures. Scores on the ASVAB could also correlate with scores on tests of executive functions such as attention control and placekeeping ability because the different instruments “sample” some of the same domain-general cognitive processes, including not only those executive functions but any number of other domain-general cognitive processes as well (see [Kovacs and Conway 2016](#)).

Regardless of whether measures of placekeeping ability and attention control explain the predictive validity of the ASVAB, they may improve the prediction of job-relevant performance above and beyond the ASVAB. In other words, ASVAB scores and executive control measures may account for unique portions of variance in job-relevant performance. If so, tests of executive functions would be promising candidates to augment the ASVAB.

2. Present Study

Our major research question was whether Gf, placekeeping ability, and attention control would explain the relationship between AFQT score and performance in “synthetic work” tasks designed to measure job-relevant multitasking performance. The goal of the synthetic work approach, developed in the 1960s by human factors psychologists ([Alluisi 1967](#); [Morgan and Alluisi 1972](#)), is to create laboratory tasks that capture requirements common to a broad class of work tasks rather than to simulate a specific job. Thus, a synthetic work task does not resemble any actual work task *by design*. The most widely administered synthetic work task was developed for the U.S. Army by [Elsmore \(1994\)](#). The participant is to manage component tasks presented in quadrants of their computer screen: a self-paced math task (upper right), a memory probe task (upper left), a visual monitoring task (lower left), and an auditory monitoring task (lower right; further details are presented below). Total points, displayed in the center of the screen, is the sum of points earned in the component tasks. The synthetic work approach helps predict performance on classes of jobs, including potential jobs of the future workplace, complementing selection instruments that measure performance on specific jobs.

We asked whether Gf, placekeeping ability, and attention control are distinct dimensions of variation in cognitive functioning that contribute to individual differences in multitasking, a performance context in which “cognitive processes involved in performing two (or more) tasks overlap in time” ([Koch et al. 2018](#), p. 558). The role of Gf in multitasking may be to develop a strategy (or plan) for sequencing the component tasks based on the payoffs of the component tasks and considerations about when it is advantageous to sacrifice performance in one component task to attend to another. The role of placekeeping ability may be to implement the strategy or plan, allowing the task performer to keep track of which component tasks have been performed and which remain to be performed. Finally, attention control may serve multiple functions in multitasking, such as maintaining focus on a given component task while suppressing information from other component tasks, disengaging from one component task to shift to another one, and monitoring performance.

The traditional approach to answering a research question such as ours is structural equation modeling (SEM). In SEM, *latent variables* are specified as “common causes” of individual differences in sets of *observed variables*, and then causal relations among the latent variables are estimated via path analysis. However, there is not always a strong justification for making causal assumptions concerning relations among variables. For example, attention control is a plausible cause of individual differences in the reasoning tests we used to measure Gf, as the ability to maintain information from a reasoning problem in working memory is presumably critical for being able to solve that problem. However, Gf is a plausible cause of individual differences in the tasks we used to measure attention control, given that those tasks were novel to participants, and given evidence that strategy use influences performance in even very simple cognitive tasks (e.g., [Deary 2000](#)).

Psychometric network analysis offers an alternative approach to investigating relations among psychological variables that does not rest on assumptions about causality (see [Schmank et al. 2019](#); [Borsboom et al. 2021](#); [McGrew 2023](#)). Each variable in the analysis is represented by a *node* (a circle).¹ The nodes are connected by non-directional *edges* (lines)

and association strength is described with partial correlations (or some other statistical parameter), with thicker edges indicating stronger associations. All possible nondirectional associations are identified, and thus the graphical representation of the results (the *network model*) captures all possible ways that any two variables may relate to each other. In a network model of cognitive ability variables where values for links are partial correlations, the link between two variables represents the association between those variables controlling for the *g* factor in the data (McGrew 2023).

Here, to answer our major research question, we generated and compared two network models in which the nodes were composite variables (represented by circles) created by averaging the observed variables corresponding to the different factors. In the *bivariate model*, association strength was indicated by bivariate correlations (Pearson's *rs*). In the *partial model*, association strength was indicated by partial correlations (*prs*), each reflecting the association between two variables statistically controlling for the other variables in the model. A finding that the bivariate correlation between AFQT and multitasking performance was significant would be consistent with existing research and practice. A finding that the corresponding partial correlation was also significant would indicate that variability in AFQT directly explains variability in multitasking performance after partialling out the variance shared with other measures collected in this study. In contrast, a finding that the partial correlation was nonsignificant may indicate the relationship between AFQT and multitasking performance was explained by one or more of the other predictor variables (*Gf*, attention control, and placekeeping ability).

An additional goal of this study was to evaluate the incremental validity of *Gf*, placekeeping ability, and attention control in the prediction of multitasking performance. In an influential series of studies, Ree, Carretta, and colleagues showed that variance specific to the ASVAB subtests ("s" factors) added little to the prediction of job (and training) performance over variance common to the subtests (e.g., Ree et al. 1994; Ree and Earles 1992). Furthermore, in factor-analytic studies, correlations between factors from the ASVAB and those representing constructs from cognitive psychology such as working memory capacity have been found to be very high (Kyllonen 1993), although less than unity (Kane et al. 2004; Ackerman et al. 2005). The view that has emerged from such evidence is that predicting job performance is "not much more than *g*" (Ree et al. 1994) as measured by the ASVAB. Recently, Ree and Carretta (2022) concluded that predicting job performance is "still not much more than *g*" (p. 1, emphasis added).

This view implies that the ASVAB measures all aspects of the cognitive system that may bear on job performance. Yet, the predictive validity of the ASVAB is far from perfect. For example, if the validity of the ASVAB across all jobs is 0.60, then well over half of the variance in job performance (64%) is explained by factors other than those measured by the ASVAB (i.e., $100 - 0.60^2 \times 100 = 64$). At least some of this unexplained variance may be explained by *Gf*, placekeeping ability, and attention control. These factors have all been found to correlate strongly with multitasking performance as measured with synthetic work tasks (Burgoyne et al. 2021b; Hambrick et al. 2011; Redick et al. 2016; Martin et al. 2020), but the present study is the first to measure all three factors and multitasking performance in the same sample of participants.

We took two approaches to testing the incremental validity of *Gf*, placekeeping ability, and attention control for multitasking. As a weaker test, we performed hierarchical regression analyses to evaluate the extent to which *Gf*, placekeeping ability, and attention control each predicted multitasking performance over AFQT. As a stronger test, we evaluated the extent to which each factor improved the prediction of multitasking performance over AFQT and the other factors we measured.

3. Method

3.1. Participants

Participants were undergraduate students from a large midwestern (U.S.) university recruited through the Department of Psychology subject pool and received credit toward

a course requirement for volunteering. The study was approved by the university’s Institutional Review Board, and all participants gave informed consent at the beginning of the study.

Participants ranged in age from 18 to 26 years ($M = 19.0$, $SD = 1.1$); 70.5% reported they were female, 26.9% reported they were male, and 2.6% did not report gender. The breakdown of self-reported ethnic/racial group was Black (11.4%), White (72.0%), Asian (10.3%), Other (3.7%), and no response (2.6%).

With respect to college admissions scores, the mean ACT score ($n = 112$) was 26 ($SD = 4$, Range = 16–35), compared to a mean of 21 for all test takers ($SD = 6$, Range = 1 to 36), and the mean SAT score ($n = 175$) was 1195 ($SD = 134$, Range = 880–1550), compared to a mean of 1059 for all test takers ($SD = 210$, Range = 400–1600; [National Center for Education Statistics 2019](#)). Thus, relative to all test takers, the mean ACT score was 0.9 SD s higher and the SD was 33% smaller, and the mean SAT score was 0.7 SD s higher and the SD was 37% smaller. Furthermore, on average, the four ASVAB subtest means were 0.5 SD s above means for a large sample of applicants ($N = 8808$) for military service ([Ree and Wegner 1990](#)). In sum, the sample was higher on average than the general population in cognitive ability, and although there was some degree of range restriction, it was not severe.

Of the 270 participants who completed Session 1, 29 did not return for Session 2 (leaving $n = 241$), another 26 did not return for Session 3 (leaving $n = 215$), and another 57 did not return for Session 4 (leaving $n = 158$). It appears attrition was random with respect to performance on the predictor and criterion tasks. The point-biserial correlations of the performance measures from each session with a binary variable indicating whether a participant showed up for the next session (0 = no, 1 = yes) were all near zero and nonsignificant (mean $r_{pb} = 0.04$). A likely reason for the markedly higher attrition from Session 3 to 4 is that after three sessions participants had earned enough experiment credits to fulfill the research participation assignment in their course.

3.2. Procedure and Materials

This study took place in four sessions, each lasting 1 to 1.5 h. Participants were tested in groups of up to 7. As is standard in individual difference research (e.g., [Martin et al. 2020](#)), the measures were collected in a fixed order to avoid participant \times order interactions. Table 2 lists the tests for each session, in the order administered.²

Table 2. Assessments by Session in Order Administered.

Session 1	Session 2	Session 3	Session 4
Consent form	Foster Task	Control Tower	SynWin
Demographic form	(MP)	(MP)	(MP)
Arithmetic Reasoning (ASVAB)	UNRAVEL (PA)	Letter Sets (Gf)	Stroop Deadline (AC)
Word Knowledge (ASVAB)	Mental Counters (WMC)	Antisaccade (AC)	Selective Visual Arrays (AC)
Mathematics Knowledge (ASVAB)	Raven’s Matrices (Gf)	Letterwheel (PA)	
Paragraph Comprehension (ASVAB)	Sustained Attention to Cue (AC)	Number Series (Gf)	
		Flanker Deadline (AC)	

Note. ASVAB, Armed Services Vocational Aptitude Battery; MP, Multitasking Performance. PA, Placekeeping Ability; WMC, Working Memory Capacity; Gf, Fluid Intelligence; AC, Attention Control.

Armed Services Vocational Aptitude Battery (ASVAB). Participants completed four subtests from a practice ASVAB. The four subtests were those comprising the Armed Forces Qualification Test (AFQT), which is used to determine whether an individual is eligible to enlist in the military.³ Each test consisted of 4-alternative multiple-choice items. The number of items per test and time limits were the same as in the actual ASVAB. The sub-

tests were presented to participants in paper-and-pencil format in binders, but participants entered their answers for each subtest in a Qualtrics form. Keeping time with a digital stopwatch, an experimenter instructed participants when to begin working on each subtest and when to stop.

The four subtests comprising the AFQT were Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, and Mathematics Knowledge (see Table 1). Arithmetic Reasoning consisted of 30 word problems, and the time limit was 36 min. Word Knowledge consisted of 35 synonym questions, and the time limit was 11 min. Paragraph Comprehension consisted of 15 questions, of which the first 12 each referred to a separate paragraph and the last three referred to the same paragraph; the time limit was 13 min. Mathematics Knowledge consisted of 25 questions to assess knowledge of basic mathematic concepts, and the time limit was 24 min. For each subtest, the score was the number correct.

Attention control (AC). There were five attention control tasks (see additional details about the tasks in [Martin et al. 2020](#), and [Draheim et al. 2021](#)).⁴ In a trial of Sustained Attention to Cue, the participant saw a fixation cross for 2 or 3 s (quasi-randomly determined, half the trials for each duration), after which a 300 ms tone was played, followed by a white circle appearing in a randomly determined location, not at the center of the screen. After the circle onset, it shrank for 1.5 s until it reached a fixed size, after which there was a variable wait time of 2, 4, 8, or 12 s (quasi-randomly determined, equal numbers of trials for each duration). After the wait time, a distractor in the form of a white asterisk blinked on and off at the center of the screen for 400 ms. After the blinking asterisk, a 3 × 3 array of letters appeared at the location of the circle cue. The central letter in the array (one of B, D, P, and R) was the target and appeared in dark gray font. The other letters in the array (two each of B, D, P, and R) appeared in black font. After 125 ms the target was masked with a # character for 1 s, after which the array offset and four response boxes appeared on the screen, each containing one of the four possible targets. The participant was to choose the box with the target letter. The measure of performance was accuracy (the proportion of correct trials).

In a trial of Antisaccade, the participant saw a fixation cross for a randomly determined amount of time between 2000 and 3000 ms, after which an alerting tone was played for 300 ms. A distractor (a flashing # character) then appeared on either the left or right side of the screen for 300 ms and was followed immediately by a target letter (Q or O) for 100 ms on the side of the screen opposite to where the distractor appeared. Finally, a mask (two # characters) appeared for 500 ms at the location of the target and the distractor. The participant was to identify the target (Q or O) by pressing the corresponding key. The measure of performance was accuracy (the proportion of 72 trials that were correct).

In a trial of Flanker Deadline, the participant saw five arrows and was to indicate which direction the middle arrow was pointing. A response deadline, announced by a loud beep, limited how long the participant was allowed to respond. Each of the 18 blocks included 12 congruent and 6 incongruent trials, presented in random order with a randomized inter-stimulus interval between 400 and 700 ms. The starting deadline was 1050 ms. Between blocks, the deadline decreased if the participant was correct on at least 15 of 18 trials (by 90 ms for Blocks 1–6 and 30 ms for Blocks 7–18) and increased if they were not (by 270 ms for Blocks 1–6 and 90 ms for Blocks 7–18). The measure of attention control was what the response deadline would be for a 19th block, multiplied by –1 so that higher values would indicate better performance.

In a trial of Selective Visual Arrays, the participant saw a display composed of blue and red rectangles in different orientations on a light gray background, with 5 or 7 rectangles of each color (10 or 14 rectangles in total). Before each trial, the participant was cued to attend to either the red or blue rectangles. After a 900 ms delay, the target array appeared for 250 ms, after which an array with only the rectangles of the target color appeared. One rectangle was circled, and the participant was to respond whether it was in the same orientation as in the initial array. The rectangle was in the same orientation on 50% of trials, and there were 48 trials per array size, for 96 trials total. The measure of attention control

was the k measure of capacity [$N \times (\% \text{ hits} + \% \text{ correct rejections} - 1)$, where N = display size], averaged across array sizes.

In a trial of the Stroop Deadline, the participant saw a color word in the congruent or incongruent color. The beginning deadline was 1230 ms; otherwise, task parameters were the same as in Flanker. The measure of attention control was what the response deadline would be for a 19th block, multiplied by -1 so higher values would indicate better performance.

Fluid intelligence (Gf). There were three Gf tests. In Raven's Advanced Progressive Matrices (Raven et al. 1998; Redick et al. 2016), each item consisted of a 3×3 array of geometric patterns, with the pattern in the lower right missing. Eight alternatives appeared below the array, and the participant's task was to click on the one that fit logically as the missing pattern. We used the 18 odd items from Raven's, and the score was the proportion correct. The time limit was 10 min.

In Letter Sets (Ekstrom et al. 1976; Redick et al. 2016), each item consisted of five sets of letters in a row, with four letters per set (e.g., FQHI DFHJ YLMH ERHT VNHQ). The participant was to click on the set that did not fit logically with the other sets. There were 20 items, and the score was the proportion correct. The time limit was 10 min.

In Number Series (Thurstone 1938; Redick et al. 2016), each item consisted of a series of numbers, followed by a blank space (e.g., 1 3 6 10 _). The participant was to infer the rule governing the changes in the digits and to select the alternative that fit logically in the blank. There were 15 items, and the score was the proportion correct. The time limit was 5 min.

Multitasking performance (MP). Three synthetic work tasks were used to assess multitasking performance (see Appendix F for screenshots). In the Foster Task (developed by Jeffrey L. Foster, Macquarie University; see Martin et al. 2020), four subtasks are displayed in quadrants of the screen. In Telling Time, the participant saw a clock and was to click on one of four buttons showing the time displayed. In Visual Monitoring, the participant was to click on a disk as quickly as possible after it began spinning. In Word Recall, a word appeared in green and then disappeared. A word then appeared in red and the participant was to click a YES button if it was the same word or a NO button if it was not. In Math, the participant was to solve two-term addition problems. In each subtask, participants were awarded 100 points for correct responses and docked 100 points for incorrect responses. Participants completed a 5 min session of the Foster task; the measure of multitasking performance was the average score across the subtasks.

In Control Tower (Redick et al. 2016), participants were to perform a primary task while concurrently performing secondary tasks. In the primary task, the participant was shown two arrays of digits, letters, and symbols, which were presented in adjacent rectangles. The participant was to search through the left array and respond to items in the right array, according to the following rules: for a digit, click on the same digit; for a letter, click on the preceding letter in the alphabet; and for a symbol, click on a corresponding symbol, clicking on a CODE BOOK button on the right side of the screen to display the symbol pairings if necessary. While performing this task, the participant was to respond to interruptions from four secondary tasks. In the Radar task, the participant was to monitor a circular display in the bottom left of the screen, clicking an INSIDE button for blips appearing inside the circle and an OUTSIDE button for those appearing outside of it. In the Airplane task, the participant was to respond to landing requests presented through headphones by clicking on a RUNWAY button to view runway availability and then clicking on a CLEAR FOR LANDING button or DENY LANDING button based on this information. In the Color task, one of three colors was flashed on the screen and the participant was to click on the appropriate "error" button (ERROR 1, ERROR 2, or ERROR 3), clicking on a PROTOCOL button on the right side of the screen to view the color-error pairings. Finally, in the Problem Solving task, the participant heard trivia questions through the headphones and was to click on one of three possible answers at the bottom of the screen. Participants completed a

10 min session. The measure of multitasking performance was the average of z scores for the primary task and distractor task scores.

In SynWin (Elsmore 1994), the participant was to manage four subtasks, each presented in a different quadrant of the screen with the total score displayed in a box in the center of the screen. In Arithmetic (upper right), the participant was to add numbers using the + and – buttons and then click “Done” to register the answer. Ten points were awarded for correct responses and 10 points were deducted for incorrect responses. In Memory Search (upper left), a set of seven letters was displayed at the beginning of the session and then disappeared; probe letters were then displayed at a regular interval, and the participant’s task was to judge whether each was from the set by clicking on a Yes or No button. Ten points were awarded for correct responses and 10 points were deducted for incorrect responses or failures to respond. In Visual Monitoring (lower left), a line representing a needle moved from right to left on a fuel gauge, and the participant was to click anywhere on the gauge before the needle reached the end of the red region. Ten points were awarded for a reset occurring when the needle was in the red region and fewer points for resets occurring earlier. Finally, in Auditory Monitoring (bottom right), the participant was to respond to a high-pitched tone by clicking on an “Alert” button, while ignoring low-pitched tones. Ten points were awarded for correct responses and 10 points were deducted for incorrect responses or failures to respond. There was a 1 min practice block, followed by four 5 min test blocks. The measure of multitasking performance was the total score.

Placekeeping ability (PA). There were two placekeeping tasks, each of which involved performing a “looping” procedure (see Appendix F for screenshots of the materials, and Altmann and Hambrick 2022 and Burgoyne et al. 2019 for further details). In each trial of UNRAVEL, the participant applied a 2-alternative forced-choice (2AFC) decision rule to a randomly generated stimulus with features including a letter, a digit, and several kinds of formatting. There were seven rules, each linked mnemonically to a letter of the word UNRAVEL as follows: (1) character underlined or in italics, (2) letter near to or far from the beginning of the alphabet, (3) character red or yellow, (4) character above or below an outline box, (5) letter a vowel or consonant, (6) digit even or odd, and (7) digit less or more than 5. For each rule, the two possible responses were the first letters of the two options (e.g., in Step 1, *u* if the letter is underlined or *i* if it is italicized). (Note that the first responses spell UNRAVEL.) The next trial began immediately after the response. On the next trial, a new stimulus appeared, and the participant was to shift to the next rule in the UNRAVEL sequence, with U following L (creating the “loop”). A placekeeping error occurred if the participant applied an incorrect rule relative to the previous trial (e.g., the R rule following the U rule). The stimulus contained no information about the participant’s location in the rule sequence, so the participant had to remember this information. For help remembering the rule sequence and responses, the participant could press a key sequence to display a help screen.

Performance was periodically interrupted by a typing task that replaced the placekeeping stimulus on the display. On each trial of the typing task, the participant typed a 14-letter “code” into a gray box. The code was made up of the candidate UNRAVEL responses presented in random order. The participant could backspace to correct mistakes and pressed the Enter/Return key to enter the code when they deemed it complete. If the entered code was incorrect, the box was cleared and the participant had to re-enter the code. There were two trials of the typing task per interruption. After the second code was entered, the participant was to resume the UNRAVEL sequence with the rule following the rule applied on the last trial before the interruption. Participants completed two blocks of UNRAVEL, with a block comprising 11 runs of 2AFC trials separated by 10 interruptions.

In each trial of Letterwheel, the participant saw a “wheel” of nine letters and was to alphabetize a set of three letters located at contiguous spatial positions. The first two letters appeared in a box at the center of the wheel as they were typed, and the participant could backspace to correct mistakes. When the third letter was typed the next trial began

immediately. On the next trial, the locations of the nine letters were randomized, and the set of positions to alphabetize shifted one position clockwise on the wheel. The set of positions continued to shift one position per trial on subsequent trials, circling the wheel repeatedly (creating the “loop”). A placekeeping error occurred when the participant typed letters from an incorrect set of positions and was scored with respect to the previous trial. For help remembering the alphabet the participant could press a key sequence to display it.

Performance was periodically interrupted by a counting task that replaced the letter wheel on the display. On each trial, 7, 8, or 9 asterisks were randomly distributed on the display within the same region occupied by the letter wheel, and the participant responded by pressing the key corresponding to the number of asterisks. If the response was incorrect, the screen flashed and the participant repeated the trial with the same asterisk display. There were five counting trials per interruption, and after the fifth, the participant was to resume alphabetizing at the correct set of positions on the wheel, namely one position clockwise from the alphabetizing trial immediately before the interruption. Participants completed two blocks of Letterwheel, with a block comprising nine runs of alphabetizing trials separated by eight interruptions.

For each placekeeping task, we created a composite performance variable capturing all aspects of performance in the task (see the Appendix of [Burgoyne et al. 2021b](#)). This composite was the average of *z* scores for baseline sequence error rate, post-interruption sequence error rate, nonsequence error rate, baseline response time on correct trials, post-interruption response time on correct trials, time to complete the training phase before the test phase, and response time for interruptions, multiplied by -1 so that higher values would indicate better performance.

3.3. Data Preparation

We prepared the data for analysis in four steps. First, we deleted values of 0 or less (points) on multitasking variables, 0 (number of problems correct) on ASVAB and Gf variables, 0 (proportion of trials correct) on attention control variables, and 1 (proportion of trials incorrect) on placekeeping variables. Our assumption was that values not meeting these criteria reflected equipment error, or participants making no effort. Second, we converted the variables to *z* scores and screened for univariate outliers, defined as a *z* score greater or less than 3.5 (i.e., $>$ or $<$ 3.5 standard deviations from the total sample mean). Third, we replaced any outlier with a value of $z = 3.5$ or -3.5 , as appropriate. Finally, we used the expectation maximization (EM) procedure in SPSS 27 to replace values missing due to attrition or technical or experimenter error (18.4% of the data) and values missing due to the above deletions (0.6% of the data). Frequencies of outliers and missing and replaced values are reported in Appendix A.

Our final data set included 270 participants, but to evaluate whether estimating missing data materially affected conclusions, we performed all major analyses reported in the *Results* section using the subsample of participants who completed all four sessions ($n = 158$). The results of these analyses are summarized below and presented in Appendix C. Data files are available on OSF (https://osf.io/z4gf3/?view_only=0e79562fe39f4b0e83294f38630c5f0d).

3.4. Composite Variables

Using the cleaned data set, we computed the AFQT score for each participant by converting the subtest scores into *z* scores, and then, following the practice used in military applications, applying the formula: $2(\text{Word Knowledge } z + \text{Paragraph Comprehension } z) + \text{Mathematics Knowledge } z + \text{Arithmetic Reasoning } z$. For the other factors, based on classifications of the tasks in the extant literature, we created composite variables by averaging *z* scores of the observed measures, as follows: Gf (Raven’s Matrices, Letter Sets, Number Series), Attention Control (Antisaccade, Sustained Attention to Cue, Flanker Deadline, Stroop Deadline, Selective Visual Arrays), Placekeeping Ability (UNRAVEL composite,

Letterwheel composite), and Multitasking Performance (Foster, Control Tower composite, Synthetic Work). See Appendix A for descriptive statistics and a full correlation matrix.

We also performed a psychometric network analysis on the individual observed variables (see Appendix B). The results show mixed evidence for the coherence of the factors we adopted from the extant literature. As expected, there was evidence that the ASVAB subtests used to compute the AFQT score measure two factors: mathematical ability (Arithmetic Reasoning and Mathematics Knowledge, $pr = 0.45$) and verbal ability (Word Knowledge and Paragraph Comprehension, $pr = 0.33$). There also was evidence for a placekeeping ability factor, as there was a significant link between UNRAVEL and Letterwheel ($pr = 0.35$), consistent with previous findings (Altmann and Hambrick 2022). By contrast, although there was a significant link between two of the three Gf variables (Raven's and Letter Sets, $pr = 0.25$), the third Gf variable (Number Series) sat apart in the network model. Furthermore, although some of the links between attention control variables were significant (e.g., Selective Visual Arrays and Sustained Attention to Cue, $pr = 0.27$), others were not (e.g., Flanker Deadline and Sustained Attention to Cue, $pr = -0.01$). This result suggests that some cognitive constructs may be less psychologically coherent than others, a point to which we return in the Discussion.

3.5. Power Analyses

For the psychometric network analyses, we conducted post-hoc power analyses (in SPSS 27) for bivariate correlations and partial correlations, assuming that three variables were partialled out in the latter. With alpha set at 0.05, a sample of 270 provided power of 0.80 for $r = pr = 0.17$, 0.90 for $r = pr = 0.20$, and 0.99 for $r = pr = 0.25$. For the hierarchical regression analyses, we conducted post-hoc power analyses for increment in variance explained (ΔR^2), assuming four variables in the model with one test variable, as in our more stringent test of incremental validity. With alpha set at 0.05, a sample size of 270 provided a power of 0.80 for $\Delta R^2 = 0.029$, 0.90 for $\Delta R^2 = 0.038$, and 0.99 for $\Delta R^2 = 0.064$.

4. Results

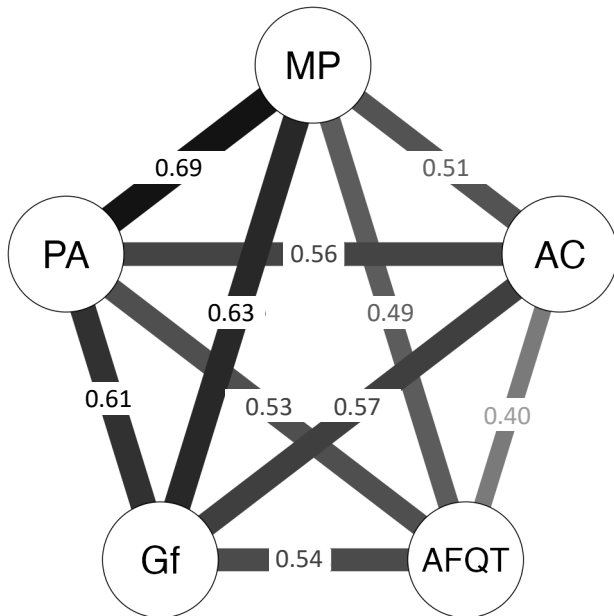
4.1. Psychometric Network Analyses

Our primary research question was whether Gf, Placekeeping Ability, and Attention Control would explain the relationship between AFQT and Multitasking Performance. To answer this question, we performed psychometric network analyses to generate two network models, presented in Figure 1. We used the Network modules in JASP Version 0.17.2; all analysis scripts are available upon request. In the first model (left panel), the values are bivariate correlations (rs). In the second model (right panel), the values are partial correlations (prs), each reflecting the association between two variables statistically controlling for their associations with the other three variables in the model. (The weight matrices for the models are presented in Table 3.) A finding that the bivariate correlation between two variables (e.g., AFQT and Multitasking Performance) is substantially larger than the partial correlation between those variables would be evidence that one or more of the other variables in the model explained the former correlation.

The AFQT–Multitasking Performance bivariate correlation ($r = 0.49$) was significant whereas the corresponding partial correlation was nonsignificant ($pr = 0.09$), indicating that one or more of the other variables in the model explained the AFQT–Multitasking Performance bivariate correlation. The bivariate correlations between AFQT and Gf ($r = 0.54$), Placekeeping Ability ($r = 0.53$), and Attention Control ($r = 0.40$) were all significant, whereas the corresponding partial correlations were significant for Gf ($pr = 0.25$) and Placekeeping Ability ($pr = 0.20$) but not Attention Control ($pr = 0.04$). This pattern indicates that the AFQT captured variance related to Gf and Placekeeping Ability but not Attention Control. The bivariate correlations between Multitasking Performance and Gf ($r = 0.63$), Placekeeping Ability ($r = 0.69$), and Attention Control ($r = 0.51$) were all significant, whereas the corresponding partial correlations were significant for Gf ($pr = 0.29$) and Placekeeping Ability ($pr = 0.42$) but not Attention Control ($pr = 0.09$). Taken together, these findings suggest that

AFQT score predicted Multitasking Performance in the bivariate model because at least some of the variability in AFQT scores was related to Gf and Placekeeping Ability.

Bivariate Correlations



Partial Correlations

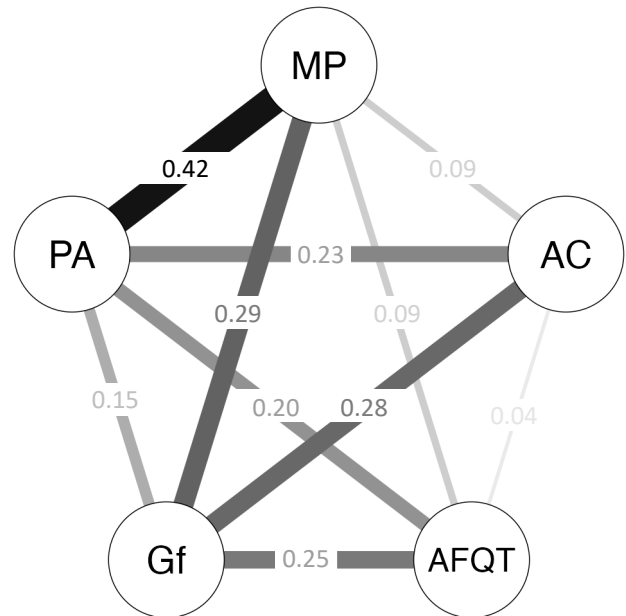


Figure 1. Psychometric network models with composites of observed variables. Values in the left panel are bivariate (Pearson’s) correlations (*rs*); values in the right panel are partial correlations (*prs*). AFQT, Armed Services Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability; AC, Attention Control; MP, Multitasking Performance. *rs* > 0.11 and *prs* > 0.12 are statistically significant (*p* < .05).

Table 3. Weight Matrices for Psychometric Network Models.

Composite Variable	1	2	3	4	5
1. AFQT	-	0.25 ** [0.14, 0.36]	0.04 [-0.10, 0.17]	0.20 ** [0.05, 0.34]	0.09 [-0.03, 0.22]
2. Gf	0.54 *** [0.45, 0.62]	-	0.28 *** [0.15, 0.40]	0.15 * [0.02, 0.28]	0.29 *** [0.18, 0.41]
3. Attention Control	0.40 *** [0.30, 0.50]	0.57 *** [0.48, 0.65]	-	0.23 ** [0.08, 0.38]	0.09 [-0.05, 0.24]
4. Placekeeping Ability	0.53 *** [0.44, 0.61]	0.61 *** [0.53, 0.68]	0.56 *** [0.47, 0.63]	-	0.42 *** [0.31, 0.53]
5. Multitasking Performance	0.49 *** [0.40, 0.58]	0.63 *** [0.56, 0.70]	0.51 *** [0.42, 0.60]	0.69 *** [0.62, 0.75]	-

Note. *N* = 270. Values below the diagonal are bivariate (Pearson’s) correlations (*rs*); values above the diagonal are partial correlations (*prs*). Values in brackets are 95% confidence intervals [lower bound, upper bound], bootstrapped for *prs*. AFQT, Armed Services Qualification Test; Gf, Fluid Intelligence. *, *p* < .05. **, *p* < .01. ***, *p* < .001.

To evaluate the strength of the evidence for our conclusion, we performed a Bayesian analysis on the partial correlation network (using the JASP Version 0.17.2 Bayesian Network module, Gaussian Graphical Model estimator). The Bayes Factor (BF_{10}) is the ratio of the likelihood of the data given the alternative hypothesis, $P(D | H_1)$, to the likelihood of the data given the null hypothesis, $P(D | H_0)$. Accordingly, a $BF_{10} > 1$ supports the alternative hypothesis, whereas a $BF_{10} < 1$ supports the null hypothesis. For example, $BF_{10} = 5$ indicates

that the data are 5 times more likely to have occurred under the alternative hypothesis than under the null hypothesis, whereas $BF_{10} = 0.20$ indicates that the data are 5 times more likely to have occurred under the null hypothesis than under the alternative hypothesis (i.e., $1/0.20 = 5$). Following convention (e.g., Lee and Wagenmakers 2014), we characterized the strength of evidence for one hypothesis over the other as “weak” for a ratio of 1 to 3, “moderate” for a ratio of 3 to 10, “strong” for a ratio of 10 to 30, and “very strong” for a ratio greater than 30.

The results are shown in Table 4 and Figure 2. For each edge (path), the alternative hypothesis is that the partial correlation has an absolute magnitude greater than zero ($H_1: pr > 0$), whereas the null hypothesis is that the partial correlation is zero ($H_0: pr = 0$). The edge inclusion criterion in Figure 2 was a $BF_{10} > 3$. There was very strong evidence for H_1 for the edges between AFQT and both Gf and Placekeeping Ability ($BF_{10}s > 100$), between Gf and both Attention Control and Multitasking Performance ($BF_{10}s > 100$), and between Attention Control and Placekeeping Ability and Placekeeping Ability and Multitasking Performance ($BF_{10}s > 100$). Furthermore, there was moderate or better evidence for H_0 for the edges between AFQT and both Attention Control ($BF_{10} = 0.06$) and Multitasking Performance ($BF_{10} = 0.12$) and between Attention Control and Multitasking Performance ($BF_{10} = 0.15$), and weak evidence for H_0 for the path between Gf and Placekeeping Ability ($BF_{10} = 0.43$).

Table 4. Bayes Factors for Partial Correlation Network Model.

Composite Variable	1	2	3	4	5
1. AFQT	-	0.00	16.67	0.00	8.33
2. Gf	∞	-	0.00	2.33	0.00
3. Attention Control	0.06	∞	-	0.00	6.67
4. Placekeeping Ability	∞	0.43	∞	-	0.00
5. Multitasking Performance	0.12	∞	0.15	∞	-

Note. $N = 270$. Values below the diagonal are BF_{10} , the odds of the data given the alternative hypothesis ($H_1: pr > 0$). Values above the diagonal are $1/BF_{10}$, the odds of the data given the null hypothesis ($H_0: pr = 0$). AFQT, Armed Services Qualification Test; Gf, Fluid Intelligence.

Partial Correlations

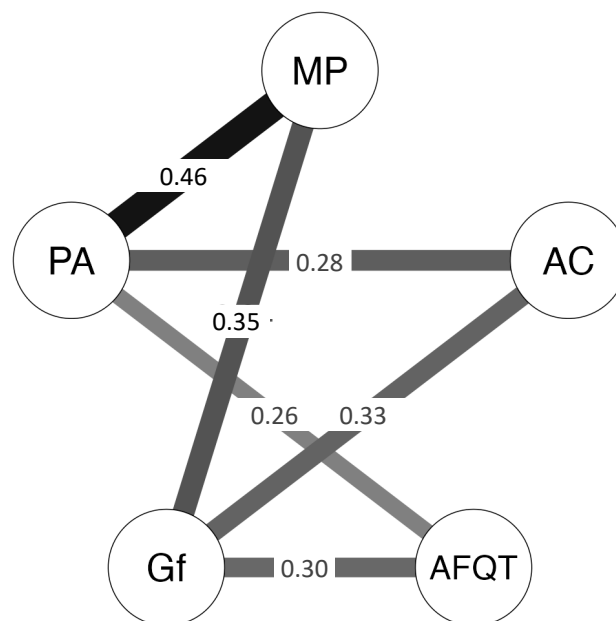


Figure 2. Bayesian psychometric network models with composites of observed variables. Values are partial correlations (prs). Inclusion criterion, $BF_{10} > 3$. AFQT, Armed Services Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability; AC, Attention Control; MP, Multitasking Performance.

Taken together, these results are compelling evidence that AFQT predicted Multitasking Performance in our sample because it indexed Gf and Placekeeping Ability, and that Attention Control played no role in the relationship between AFQT and Multitasking Performance. We repeated the psychometric network analyses using the subsample of participants who completed all sessions; the conclusions were the same as with the full sample (see Appendix C, Figure A3).

4.2. Hierarchical Regression Analyses

We performed hierarchical regression analyses to test for the incremental validity of Gf, Placekeeping Ability, and Attention Control in the prediction of Multitasking Performance over and above the AFQT. The results are presented in Tables 5–7, including the change in variance explained (ΔR^2) for the steps of each model, along with the standardized regression coefficient (β) for each predictor variable entered in those steps, reflecting the predicted increase in Multitasking Performance in SD units for every 1 SD increase in the predictor, controlling for the other predictor(s) in the model.

Table 5. Hierarchical Regression Analyses Predicting Multitasking Performance (Model 1: Each Cognitive Factor Over AFQT).

	ΔR^2	ΔF	<i>df</i>	β [95% CI]	<i>t</i>
Model 1A					
Step 1	0.241	85.17 ***	1, 268		
AFQT				0.49 [0.39, 0.60]	9.23 ***
Step 2	0.193	91.06 ***	1, 267		
AFQT				0.21 [0.10, 0.32]	3.83 ***
Gf				0.52 [0.41, 0.63]	9.54 ***
Model 1B					
Step 1	0.241	85.17 ***	1, 268		
AFQT				0.49 [0.39, 0.60]	9.23 ***
Step 2	0.257	136.62 ***	1, 267		
AFQT				0.18 [0.08, 0.28]	3.45 **
PA				0.60 [0.50, 0.70]	11.69 ***
Model 1C					
Step 1	0.241	85.17 ***	1, 268		
AFQT				0.49 [0.39, 0.60]	9.23 ***
Step 2	0.119	49.54 ***	1, 267		
AFQT				0.34 [0.23, 0.44]	6.33 ***
AC				0.38 [0.27, 0.48]	7.04 ***

Note. *N* = 270. AFQT, Armed Services Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability; AC, Attention Control. ΔR^2 , change in variance explained. β , standardized regression coefficient. **, *p* < .01. ***, *p* < .001.

In the first set of models (Table 5), we tested the incremental validity of each cognitive factor over AFQT. AFQT explained 24.1% of the variance ($\Delta F = 85.17, p < .001, \beta = 0.49$). Gf explained 19.3% of the variance over AFQT (Model 1A, $\Delta F = 91.06, p < .001, \beta = 0.52$). Placekeeping Ability explained 25.7% of the variance over AFQT (Model 1B, $\Delta F = 136.62, p < .001, \beta = 0.60$). Attention Control explained 11.9% of the variance over AFQT (Model 1C, $\Delta F = 49.54, p < .001, \beta = 0.38$).

In the second set of models (Table 6), we tested the incremental validity of each cognitive factor over AFQT and each other cognitive factor. Gf explained 5.5% of the variance over AFQT and Placekeeping Ability (Model 2A, $\Delta F = 32.46, p < .001, \beta = 0.31$), and 10.0% of the variance over AFQT and Attention Control (Model 2B, $\Delta F = 49.42, p < .001, \beta = 0.42$). Placekeeping Ability explained 11.8% of the variance over AFQT and Gf (Model 2C, $\Delta F = 70.43, p < .001, \beta = 0.46$), and 15.6% of the variance over AFQT and Attention Control (Model 2D, $\Delta F = 85.87, p < .001, \beta = 0.52$). Attention Control explained 2.6% of the variance over AFQT and Gf (Model 2E, $\Delta F = 12.84, p < .001, \beta = 0.20$), and 1.8% of the variance over AFQT and Placekeeping Ability (Model 2F, $\Delta F = 9.95, p < .001, \beta = 0.16$).

Table 6. Hierarchical Regression Analyses Predicting Multitasking Performance (Model 2: Each Cognitive Factor Over AFQT and Each Other Cognitive Factor).

	ΔR^2	ΔF	<i>df</i>	β [95% CI]	<i>t</i>
Model 2A					
Step 1	0.498	132.44 ***	2, 267		
AFQT				0.18 [0.08, 0.28]	3.45 **
PA				0.60 [0.50, 0.70]	11.69 ***
Step 2	0.055	32.46 ***	1, 266		
AFQT				0.08 [−0.02, 0.18]	1.62
PA				0.46 [0.35, 0.56]	8.39 ***
Gf				0.31 [0.20, 0.42]	5.70 ***
Model 2B					
Step 1	0.360	75.06 ***	2, 267		
AFQT				0.34 [0.23, 0.44]	6.33 ***
AC				0.38 [0.27, 0.48]	7.04 ***
Step 2	0.100	49.42 ***	1, 266		
AFQT				0.18 [0.08, 0.29]	3.38 **
AC				0.20 [0.09, 0.31]	3.58 ***
Gf				0.42 [0.31, 0.54]	7.03 ***
Model 2C					
Step 1	0.434	102.42 ***	2, 267		
AFQT				0.21 [0.10, 0.32]	3.83 ***
Gf				0.52 [0.41, 0.63]	9.54 ***
Step 2	0.118	70.43 ***	1, 266		
AFQT				0.08 [−0.02, 0.18]	1.62
Gf				0.31 [0.20, 0.42]	5.70 ***
PA				0.46 [0.35, 0.56]	8.39 ***
Model 2D					
Step 1	0.360	75.06 ***	2, 267		
AFQT				0.34 [0.23, 0.44]	6.33 ***
AC				0.38 [0.27, 0.48]	7.04 ***
Step 2	0.156	85.87 ***	1, 266		
AFQT				0.15 [0.05, 0.25]	2.97 **
AC				0.16 [0.06, 0.27]	3.15 **
PA				0.52 [0.41, 0.63]	9.27 ***
Model 2E					
Step 1	0.434	102.42 ***	2, 267		
AFQT				0.21 [0.10, 0.32]	3.83 ***
Gf				0.52 [0.41, 0.63]	9.54 ***
Step 2	0.026	12.84 ***	1, 266		
AFQT				0.18 [0.08, 0.29]	3.38 **
Gf				0.42 [0.31, 0.54]	7.03 ***
AC				0.20 [0.09, 0.31]	3.58 ***
Model 2F					
Step 1	0.498	132.44 ***	2, 267		
AFQT				0.18 [0.08, 0.28]	3.45 **
PA				0.60 [0.50, 0.70]	11.69 ***
Step 2	0.018	9.95 ***	1, 266		
AFQT				0.15 [0.05, 0.25]	2.97 **
PA				0.52 [0.41, 0.63]	9.27 ***
AC				0.16 [0.06, 0.27]	3.15 **

Note. *N* = 270. AFQT, Armed Services Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability; AC, Attention Control. ΔR^2 , change in variance explained. β , standardized regression coefficient. **, *p* < .01. ***, *p* < .001.

In the third set of models (Table 7), we tested the incremental validity of each cognitive factor over AFQT and both other cognitive factors. Gf explained 4.0% of the variance over AFQT, Placekeeping Ability, and Attention Control (Model 3A, $\Delta F = 24.17$, *p* < .001, $\beta = 0.28$). Placekeeping Ability explained 9.6% of the variance over AFQT, Gf, and Attention Control (Model 3B, $\Delta F = 57.59$, *p* < .001, $\beta = 0.43$). Attention Control explained a nonsignificant 0.4%

of the variance over AFQT, Gf, and Placekeeping Ability (Model 3C, $\Delta F = 2.36$, $p = .125$, $\beta = 0.08$).

Table 7. Hierarchical Regression Analyses Predicting Multitasking Performance (Model 3: Each Cognitive Factor Over AFQT and Both Other Cognitive Factors).

	ΔR^2	ΔF	<i>df</i>	β [95% CI]	<i>t</i>
Model 3A					
Step 1	0.516	94.57 ***	3, 266		
AFQT				0.15 [0.05, 0.25]	2.97 **
PA				0.52 [0.41, 0.63]	9.27 ***
AC				0.16 [0.06, 0.27]	3.15 **
Step 2	0.040	24.17 ***	1, 265		
AFQT				0.08 [−0.02, 0.18]	1.54
PA				0.43 [0.32, 0.54]	7.59 ***
AC				0.08 [−0.02, 0.19]	1.54
Gf				0.28 [0.17, 0.40]	4.92 ***
Model 3B					
Step 1	0.460	75.59 ***	3, 266		
AFQT				0.18 [0.08, 0.29]	3.38 **
Gf				0.42 [0.31, 0.54]	7.03 ***
AC				0.20 [0.09, 0.31]	3.58 ***
Step 2	0.096	57.59 ***	1, 265		
AFQT				0.08 [−0.02, 0.18]	1.54
Gf				0.28 [0.17, 0.40]	4.92 ***
AC				0.08 [−0.02, 0.19]	1.54
PA				0.43 [0.32, 0.54]	7.59 ***
Model 3C					
Step 1	0.553	109.51 ***	3, 266		
AFQT				0.08 [−0.02, 0.18]	1.62
Gf				0.31 [0.20, 0.42]	5.70 ***
PA				0.46 [0.35, 0.56]	8.39 ***
Step 2	0.004	2.36	1, 265		
AFQT				0.08 [−0.02, 0.18]	1.54
Gf				0.28 [0.17, 0.40]	4.92 ***
PA				0.43 [0.32, 0.54]	7.59 ***
AC				0.08 [−0.02, 0.19]	1.54

Note. *N* = 270. AFQT, Armed Services Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability; AC, Attention Control. ΔR^2 , change in variance explained. β , standardized regression coefficient. **, $p < .01$. ***, $p < .001$.

To sum up, Gf, Placekeeping Ability, and Attention Control each improved the prediction of Multitasking Performance over AFQT. Moreover, each factor improved the prediction of Multitasking Performance over AFQT and each other factor, although the increments in variance explained were notably higher for Gf and Placekeeping Ability than for Attention Control. Finally, Gf and Placekeeping Ability each improved the prediction of Multitasking Performance over AFQT and both other factors, whereas Attention Control did not. Thus, overall, evidence for incremental validity was stronger for Gf and Placekeeping Ability than for Attention Control.

In Appendix C we repeat the hierarchical regression analyses using the subsample of participants who completed all sessions; the conclusions were the same as with the full sample. In Appendix E we report simultaneous regressions of Multitasking Performance on all four predictors individually and in combination, to support an assessment of the amount of criterion variance explained as a function of number of predictors.

5. Discussion

The development of reliable and valid tests of cognitive ability for the prediction of job performance (among other uses) is arguably one of applied psychology’s greatest achievements. The ASVAB, and more specifically the AFQT score, is a compelling case in

point. However, it is unclear what cognitive factors explain the predictive validity of the ASVAB. In this study, we investigated the contributions of three cognitive abilities—Gf, placekeeping ability, and attention control—to the relationship between the AFQT score and job-relevant multitasking performance.

We compared two psychometric network analyses of relations among composite variables representing the predictor and criterion (see Figure 1). The correlation between AFQT and multitasking performance was strongly positive, but the corresponding partial correlation was near zero. Moreover, although the correlations of both AFQT and multitasking performance with the other predictor variables were strongly positive, the corresponding partial correlations were significant for Gf and placekeeping ability but not attention control. The results are evidence that Gf and placekeeping ability contributed independently to the relationship between AFQT and multitasking performance, whereas attention control did not.

These findings are consistent with the possibility that Gf and placekeeping ability are causal factors underlying individual differences in multitasking. Gf measures the ability to develop solutions to novel problems, and placekeeping measures the ability to maintain one's place in a sequence of mental operations without omissions or repetitions. Thus, one interpretation of the independent contributions of the two factors to multitasking performance is that Gf measures success at developing effective multitasking strategies, and placekeeping ability measures success at executing those strategies once they are developed. The significant partial correlations of attention control with Gf and placekeeping suggest that attention is a primitive cognitive operation underlying both (Burgoyne and Engle 2020).

Following up on a study by Martin et al. (2020), we evaluated the incremental validity of Gf, placekeeping ability, and attention control in the prediction of multitasking performance. Novel in our study were the additions of Gf and placekeeping ability. Each of Gf, placekeeping ability, and attention control improved the prediction of multitasking performance over AFQT, but only Gf and placekeeping ability did so over all other predictor variables (i.e., AFQT and, respectively, placekeeping ability and Gf). These results indicate that tests of Gf and placekeeping ability are perhaps the most promising candidates for augmenting the ASVAB to predict military job performance, at least in military occupation specialties that place high demands on multitasking. Appendix E presents additional information about how the ASVAB might be augmented, in the form of R^2 values for each predictor variable alone and in all possible combinations.

The effect sizes for attention control in the present study were smaller than those in Martin et al. (2020). In particular, Martin et al. found that three attention control measures (Antisaccade, Sustained Attention-to-Cue, and Selective Visual Arrays) explained 6.4% of the variance in multitasking performance over ASVAB and Gf. By contrast, in the present study, the attention control composite explained 2.6% of the variance over ASVAB and Gf (see Table 6, Model 2E). This difference does not appear to be due to our measures of attention control having poor psychometric properties. Reliability estimates for our measures of attention control, reported in Table A2, were similar in magnitude to those reported by Martin et al. (2020). Furthermore, indicative of construct validity, the five measures of attention control correlated moderately and positively with each other, with an average correlation of 0.40 (range = 0.27–0.56; see Table A2). The average correlation among the three measures of attention control that Martin et al. used in their main analyses was very similar (avg. $r = 0.43$; range = 0.33–0.47). The weaker effect sizes in our study may instead reflect a greater range of cognitive ability in the Martin et al. sample than in ours, given that their sample included both undergraduate students and community-recruited non-students, whereas ours included only undergraduate students. In fact, there was greater variability in some of the attention control measures in Martin et al.'s study than in our study (e.g., Antisaccade SDs of 0.15 and 0.11, respectively). Finally, the difference could also be due to sampling error, as the sample sizes in Martin et al.'s study ($N = 171$) and this study ($N = 270$) were typical for this area of research though still somewhat small.

Concerning our finding of stronger incremental validity for placekeeping ability than for attention control in the prediction of multitasking performance, we offer two explanations that are not mutually exclusive. First, placekeeping ability constitutes a complex suite of cognitive processes, whereas attention control represents a set of more basic cognitive operations. Thus, placekeeping may have been relatively closer to, and thus accounted for more variance in, the outcome variable. Second, our composite measure of performance on each placekeeping task (UNRAVEL and Letterwheel) comprised seven separate measures that independently and exhaustively captured all aspects of performance on that task (as described in the Appendix of [Burgoyne et al. 2021b](#)). By contrast, for attention control, there was only a single measure of performance for each task. Thus, placekeeping ability may have explained more variance in the criterion variable because our placekeeping measures captured more aspects of performance. One might argue that this difference in the scope of the measures made for an unfair comparison of the two constructs in terms of incremental validity. However, from both a theoretical and a practical perspective, we could think of no particular reason to exclude any of the performance measures from the placekeeping tasks. We further note that it would likely be profitable to take this “whole task” approach in the measurement of attention control, among other executive functions. For example, a composite measure of performance on the antisaccade task might include the time to proceed through instructions that require performing practice trials correctly. A study comparing incremental validity of placekeeping ability, attention control, and other factors using a whole-task approach may be another useful avenue for future research.

6. Limitations

We note four limitations of this study. First, a psychometric network analysis of the individual observed variables (see Appendix B) suggested that some of the cognitive constructs we considered in this study are more coherent than others, at least as measured by the tasks we used in this study. Most notably, while there was a strong and statistically significant partial correlation between UNRAVEL and Letterwheel ($pr = 0.35$), only half of the links between the attention control variables (5 of 10) were significant. This suggests that the attention control tasks we used in this study may measure different aspects of attention control and/or other cognitive abilities. It is also possible that some of the links between attention control variables reflect method variance, especially the link between the two tasks that used a response deadline procedure (Stroop Deadline and Flanker Deadline, $pr = 0.36$). Note that some of the attention control tasks we used in this study have been refined and other attention control tasks have been introduced (see the “Squared” tasks of [Burgoyne et al. 2023](#)). In future studies, we hope to examine whether results using these new attention control tasks are different than those reported here, especially with regard to predictive validity. Overall, these results suggest that network analysis could be a useful converging operation to evaluate the coherence of putative psychological constructs.

A second limitation is that we focused on only one aspect of job-relevant performance—multitasking—and measured that using the synthetic work approach. Multitasking is a requirement of many occupations in the military and civilian workplace ([National Center for O*NET Development 2023](#)), but the evidence presented here is no guarantee that Gf and placekeeping ability will significantly predict performance in any particular job as a whole (e.g., air-traffic control). Thus, it will be critical to evaluate whether the factors we tested predict performance in actual jobs above and beyond the ASVAB score. From our perspective, it would be ideal for studies to collect both measures of synthetic work performance and measures of real-world work performance.

A third limitation is the representativeness of the sample. A well-known problem with using “samples of convenience” in psychological research (viz., undergraduate students) is that the participants may not accurately reflect the population of interest. As already mentioned, the problem is acute in studies of cognitive ability because undergraduate students have typically already been selected on cognitive ability (i.e., score on a college

admissions test). As indicated by college admissions test scores and comparison of ASVAB subtest means to those for a military sample, there was a relatively wide range of cognitive ability in our sample. However, there was some degree of range restriction in our sample (as reflected in higher means and smaller *SDs* for the sample on college admissions tests relative to all test takers). Thus, the effect sizes reported here may be underestimates of relationships that would be observed in the general population.

One approach to addressing this limitation might be online testing, which allows for larger and more representative samples to be collected than are practical to obtain with in-person studies. Online testing may also make it easier to test military personnel (or civilian employees) and would make it easier to obtain stratified samples to investigate questions such as whether racial/ethnic group differences are smaller in measures of executive functions than those found on standardized cognitive ability tests (Burgoyne et al. 2021a). The testing environment of in-person studies is obviously different than that of online studies, beginning with the presence of an experimenter in the former but not the latter. Therefore, it would be critical to validate online tasks before drawing conclusions from them.

Finally, it is possible that some broad, unmeasured cognitive factor explains the effects of both *Gf* and placekeeping ability on multitasking performance. Processing speed could be one such factor, although it is generally found to explain quite small amounts of inter-individual variability in complex task performance (e.g., Conway et al. 2002). A more likely possibility is working memory capacity (WMC). Historically, WMC has been closely associated with *Gf* (e.g., Kyllonen and Christal 1990; Kane et al. 2004), and may also be associated with placekeeping ability (but see Burgoyne et al. 2019). Especially relevant, in a study of applicants for air-traffic control training, Colom et al. (2010) found that *Gf* added negligibly to the prediction of multitasking performance, as assessed with laboratory tasks, independent of WMC. As we noted earlier, we did collect a single measure of WMC (Mental Counters), and when we included it as a variable in our psychometric network analyses the results were largely unchanged and the conclusions were the same (see Appendix D). Moreover, whereas the bivariate correlations of WMC with AFQT ($r = 0.38$) and multitasking performance ($r = 0.54$) were significant, the corresponding partial correlations were nonsignificant ($pr_s = -0.06$ and 0.05 , respectively). Thus, the data we have suggests that WMC may not contribute to the relationship between AFQT and multitasking performance independent of *Gf*, placekeeping ability, and attention control. At the same time, this result may be specific to the Mental Counters measure, which does not always correlate highly with other measures of WMC (e.g., Jastrzębski et al. 2023). Replication with a composite variable based on multiple measures of WMC may be a useful avenue for future work.

7. Conclusions

Psychometric network analyses revealed that measures of two cognitive abilities—*Gf* and placekeeping ability—largely explained the relationship between AFQT score from the ASVAB and job-relevant multitasking performance. From a theoretical perspective, this evidence suggests that high levels of multitasking performance can be understood in terms of success in developing multitasking strategies (as measured by *Gf*) and success in implementing strategies during performance (as measured by placekeeping ability). From an applied perspective, our results suggest new directions for the development of cognitive test batteries that are even more predictive of job performance than those currently available. Finally, from a methodological perspective, the study illustrates the usefulness of psychometric network analysis in cognitive research. This approach could be extended to investigating the explanatory relationship between cognitive ability and societally relevant outcomes such as academic achievement, health, and mortality (Hunt 2010).

Author Contributions: Conceptualization, D.Z.H., A.P.B. and E.M.A.; methodology, D.Z.H., A.P.B. and E.M.A.; software, A.P.B. and E.M.A.; formal analysis, D.Z.H. and A.P.B.; resources, D.Z.H., A.P.B. and E.M.A.; data curation, D.Z.H., A.P.B. and E.M.A.; writing—original draft preparation, D.Z.H.; writing—review and editing, D.Z.H., A.P.B., E.M.A. and T.J.M.; supervision, A.P.B. and T.J.M.; project

Table A2. Cont.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
6. Sustained AC	0.22 ***	0.34 ***	0.23 ***	0.13 *	0.39 ***	(0.87)								
7. Flanker Deadline	0.34 ***	0.38 ***	0.40 ***	0.35 ***	0.56 ***	0.31 ***	-							
8. Stroop Deadline	0.19 **	0.31 ***	0.24 ***	0.25 ***	0.38 ***	0.27 ***	0.51 ***	-						
9. Selective VA	0.45 ***	0.43 ***	0.35 ***	0.40 ***	0.37 ***	0.40 ***	0.37 ***	0.39 ***	(0.73)					
10. UNRAVEL	0.45 ***	0.33 ***	0.39 ***	0.39 ***	0.44 ***	0.26 ***	0.34 ***	0.16 **	0.31 ***	(0.73)				
11. Letterwheel	0.51 ***	0.40 ***	0.54 ***	0.57 ***	0.54 ***	0.34 ***	0.51 ***	0.27 ***	0.44 ***	0.66 ***	(0.72)			
12. Foster Task	0.16 **	0.16 **	0.33 ***	0.23 ***	0.14 *	0.06	0.15 *	0.17 **	0.24 ***	0.21 ***	0.32 ***	(0.52)		
13. Control Tower	0.44 ***	0.33 ***	0.50 ***	0.52 ***	0.47 ***	0.25 ***	0.43 ***	0.20 ***	0.27 ***	0.54 ***	0.63 ***	0.32 ***	(0.58)	
14. SynWin	0.49 ***	0.34 ***	0.42 ***	0.49 ***	0.42 ***	0.22 ***	0.39 ***	0.26 ***	0.45 ***	0.48 ***	0.62 ***	0.23 ***	0.39 ***	(0.77)
Multiple R	0.69	0.63	0.67	0.69	0.69	0.54	0.70	0.58	0.66	0.70	0.83	0.44	0.73	0.69

Note. $N = 270$. AFQT, Armed Forces Qualification Test. AC, Attention to Cue. VA, Visual Arrays. Values in parentheses along the diagonal are internal consistency reliability estimates; multiple R s are provided as another lower-bound estimate of reliability. There is not yet an accepted procedure for computing internal consistency reliability for Flanker Deadline and Stroop Deadline, but both measures correlated moderately with each other ($r = 0.51$), and multiple R s were reasonably high for both measures (0.70 and 0.58, respectively), indicating acceptable reliability. *, $p < .05$. **, $p < .01$. ***, $p < .001$.

Appendix B

Figure A1 presents a network model containing all individual variables (see Table A3 for the weights matrix). Each link is a partial correlation (pr), representing the relationship between two variables (nodes) controlling for all other variables in the model. Thus, each link reflects the association between two variables controlling for the g factor in the data (see McGrew 2023). Variables hypothesized to reflect the same factor should be connected by a positive and statistically significant link and should be relatively close in the network model.

As noted in the main text, the results of this analysis suggest that, as measured by tasks available to us, the constructs we considered in this study vary in coherence. The ASVAB subtests measure two coherent factors (mathematical ability and verbal ability), and the UNRAVEL and Letterwheel tasks appear to measure a coherent placekeeping ability factor, converging with previous results (Altmann and Hambrick 2022). By contrast, for Gf , two of the tests appear to measure a coherent reasoning ability factor (Raven’s Matrices and Letter Sets) while the third (Number Series) appears to measure some other factor(s). Finally, for attention control, no more than two of the five variables are significantly related. On an exploratory basis, to improve the interpretability of the results, we repeated the main psychometric network analysis (see Figure 1, main text) using a Gf observed-variable composite based on Raven’s Matrices and Letter Sets (and excluding Number Series) and an attention control composite based on the three most highly inter-related measures of this factor (Antisaccade, Flanker Deadline, and Stroop Deadline; avg. $pr = 0.25$). As shown in Figure A2, the results are very similar to those with the original observed-variable composites, and our conclusions are unchanged.

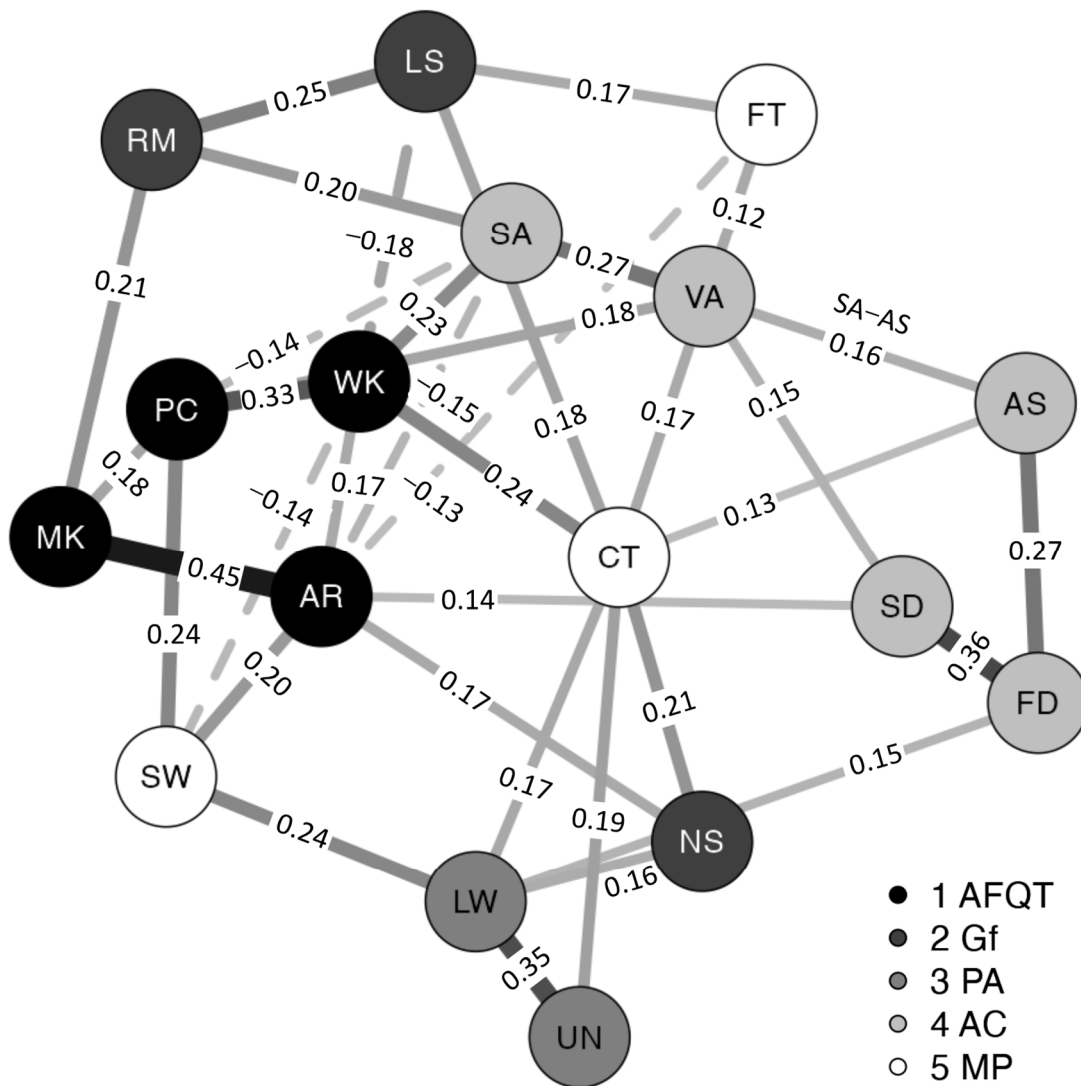


Figure A1. Psychometric network model with individual observed variables. Values are statistically significant partial correlations ($p < .05$). (1) Armed Forces Qualifying Test (AFQT) measures: Word Knowledge (WK), Paragraph Comprehension (PC), Arithmetic Reasoning (AR), and Math Knowledge (MK). (2) Fluid Intelligence (Gf) measures: Raven’s Matrices (RM), Letter Sets (LS), and Number Series (NS). (3) Placekeeping Ability (PA): UNRAVEL (UN) and Letterwheel (LW). (4) Attention Control (AC): Sustained Attention to Cue (SA), Antisaccade (AS), Flanker Deadline (FD), Stroop Deadline (SD), and Selective Visual Arrays (VA). (5) Multitasking Performance (MP) measures: Control Tower (CT), Foster Task (FT), SynWin (SW).

Table A3. Full Weights Matrix (Partial Correlations) for Psychometric Network Analysis of Individual Variables.

Variable	Network																
	RM	LS	NS	UN	LW	AS	SA	FD	SD	VA	AR	WK	MK	PC	CT	SW	FT
RM	0.000	0.249	0.021	0.054	-0.053	-0.071	0.201	0.089	0.069	0.091	0.063	0.010	0.209	0.070	-0.009	-0.040	-0.038
LS	0.249	0.000	0.064	-0.010	0.120	0.093	0.003	0.065	-0.066	0.009	0.055	-0.180	0.038	-0.076	0.182	0.061	0.167
NS	0.021	0.064	0.000	-0.080	0.163	0.014	-0.122	-0.028	0.034	0.100	0.169	0.034	0.102	0.005	0.209	0.101	-0.023
UN	0.054	-0.010	-0.080	0.000	0.348	0.116	0.001	-0.072	-0.039	-0.022	-0.028	0.053	0.078	0.059	0.186	0.100	-0.037
LW	-0.053	0.120	0.163	0.348	0.000	0.098	0.117	0.148	-0.066	0.039	0.055	0.008	0.023	0.013	0.170	0.235	0.122
AS	-0.071	0.093	0.014	0.116	0.098	0.000	0.161	0.268	0.122	0.078	-0.079	0.000	-0.023	-0.084	0.135	0.120	-0.095
SA	0.201	0.003	-0.122	0.001	0.117	0.161	0.000	-0.014	0.085	0.268	-0.147	0.230	0.003	-0.138	0.005	0.022	-0.079
FD	0.089	0.065	-0.028	-0.072	0.148	0.268	-0.014	0.000	0.360	0.003	-0.087	0.026	0.046	0.081	0.114	0.033	-0.091
SD	0.069	-0.066	0.034	-0.039	-0.066	0.122	0.085	0.360	0.000	0.152	0.137	-0.113	-0.014	-0.055	-0.051	-0.017	0.117
VA	0.091	0.009	0.100	-0.022	0.039	0.078	0.268	0.003	0.152	0.000	0.090	-0.059	0.055	0.177	-0.114	0.078	0.123
AR	0.063	0.055	0.169	-0.028	0.055	-0.079	-0.147	-0.087	0.137	0.090	0.000	0.168	0.449	0.021	0.045	0.203	-0.131
WK	0.010	-0.180	0.034	0.053	0.008	0.00	0.230	0.026	-0.113	-0.059	0.168	0.000	-0.030	0.325	0.237	-0.138	-0.029
MK	0.209	0.038	0.102	0.078	0.023	-0.023	0.003	0.046	-0.014	0.055	0.449	-0.030	0.000	0.179	0.005	-0.068	0.077
PC	0.070	-0.076	0.005	0.059	0.013	-0.084	-0.138	0.081	-0.055	0.177	0.021	0.325	0.179	0.000	-0.086	0.238	0.014
CT	-0.009	0.182	0.209	0.186	0.170	0.135	0.005	0.114	-0.051	-0.114	0.045	0.237	0.005	-0.086	0.000	-0.086	0.169
SW	-0.040	0.061	0.101	0.100	0.235	0.120	0.022	0.033	-0.017	0.078	0.203	-0.138	-0.068	0.238	-0.086	0.000	0.040
FT	-0.038	0.167	-0.023	-0.037	0.122	-0.095	-0.079	-0.091	0.117	0.123	-0.131	-0.029	0.077	0.014	0.169	0.040	0.000

Note. $N = 270$. RM, Raven’s Matrices; LS, Letter Sets; NS, Number Series; UN, UNRAVEL; LW, Letterwheel; AS, Antisaccade; SA, Sustained Attention to Cue; FD, Flanker Deadline; SD, Stroop Deadline; VA, Selective Visual Arrays; AR, Arithmetic Reasoning; WK, Word Knowledge; MK, Mathematics Knowledge; PC, Paragraph Comprehension; CT, Control Tower; SW, SynWin; FT, Foster Task. Bolded values (*prs*) are statistically significant ($p < .05$).

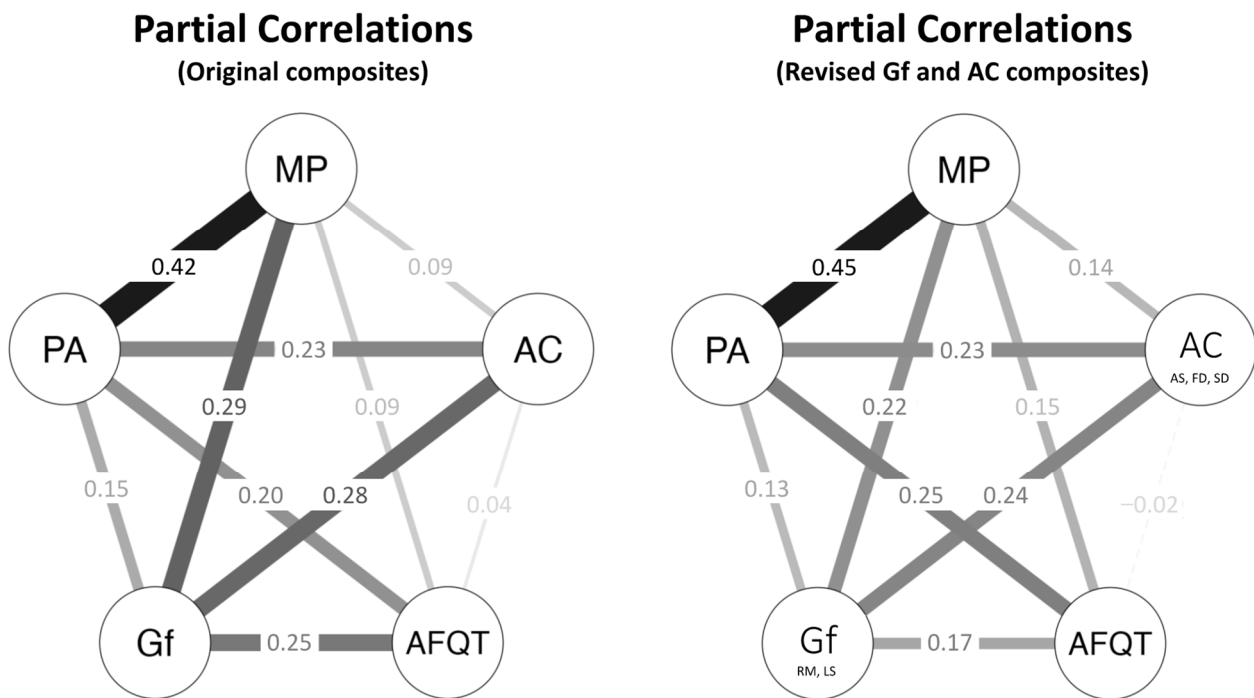


Figure A2. Psychometric network models with original composites of observed variables (left) and revised composites (right) for Gf (RM, Raven’s Matrices; LS, Letter Sets) and Attention Control (AS, Antisaccade; FD, Flanker Deadline; SD, Stroop Deadline).

Appendix C

To check whether estimating missing data materially affected our results, we reran our major analyses using the subsample of participants who completed all four sessions ($n = 158$). In this subsample, the percentage of missing data (2%) was much smaller than in the full sample (19%). To preview, the conclusions from this subsample were the same as from the full sample.

Appendix C.1. Psychometric Network Analyses

Figure A3 shows the psychometric network analysis for the subsample. The AFQT and Multitasking Performance partial correlation was non-significant ($pr = 0.12$). Moreover, the partial correlations were significant for Gf and Multitasking Performance ($pr = 0.27$), and Placekeeping Ability and Multitasking Performance ($pr = 0.38$), but not for Attention Control and Multitasking Performance ($pr = 0.11$). Thus, as in the full sample, Gf and Placekeeping Ability contributed to the AFQT and Multitasking Performance relationship, whereas Attention Control did not.

Appendix C.2. Hierarchical Regression Analyses

Table A4 presents a comparison of the incremental validities from Tables 5–7 for the full sample ($N = 270$) and the subsample of participants with all sessions ($n = 158$). As can be seen, the results are very similar, indicating that missing values had no material impact on our findings. Thus, our conclusions are unchanged.

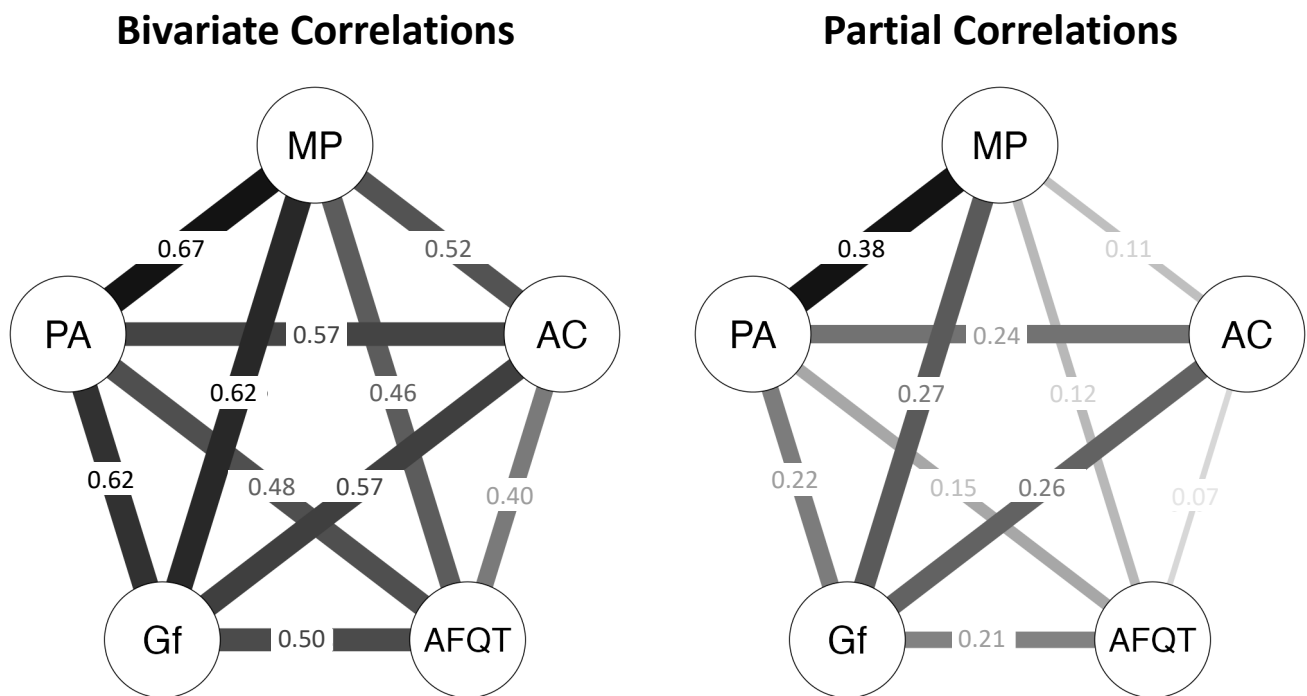


Figure A3. Psychometric network models with composites of observed variables with the subsample of participants who completed all sessions ($n = 158$). Values in the left panel are bivariate (Pearson’s) correlations (r_s); values in the right panel are partial correlations (p_r_s). MP, Multitasking Performance; AC, Attention Control; AFQT, Armed Services Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability. $r_s > 0.16$ and $p_r_s > 0.16$ are statistically significant ($p < .05$).

Table A4. Comparison of Incremental Validities Predicting Multitasking Performance for the Full Sample and the All-Sessions Subsample.

Model	Incremental Validity Test	Full Sample ($N = 270$)		All Sessions ($n = 158$)	
		ΔR^2	ΔF	ΔR^2	ΔF
1A	Gf over AFQT	0.193	91.06 ***	0.205	54.87 ***
1B	PA over AFQT	0.257	136.62 ***	0.255	74.72 ***
1C	AC over AFQT	0.119	49.54 ***	0.130	30.75 ***
2A	Gf over AFQT & PA	0.055	32.46 ***	0.053	17.03 ***
2B	Gf over AFQT & AC	0.100	49.42 ***	0.103	28.91 ***
2C	PA over AFQT & Gf	0.118	70.43 ***	0.103	33.21 ***
2D	PA over AFQT & AC	0.156	85.87 ***	0.145	43.89 ***
2E	AC over AFQT & Gf	0.026	12.84 **	0.028	7.90 **
2F	AC over AFQT & PA	0.018	9.95 **	0.020	6.02 *
3A	Gf over AFQT, PA, & AC	0.040	24.17 ***	0.053	17.03 ***
3B	PA over AFQT, Gf, & AC	0.096	57.59 ***	0.080	26.02 ***
3C	AC over AFQT, Gf, & PA	0.004	2.36	0.006	1.81

Note. $N = 270$. AFQT, Armed Forces Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability; AC, Attention Control. *, $p < .05$. **, $p < .01$. ***, $p < .001$.

Appendix D

Here we report analyses that include a measure of working memory capacity (WMC). We collected a single measure of WMC, the Mental Counters task, but omitted WMC from our main analyses because we only had the one indicator. To preview, the relationships of interest between the constructs in the main analyses replicated here.

In the Mental Counters task (Larson and Alderton 1990), a trial begins with a screen prompting the participant to imagine the digit 555. On each subsequent screen, three blank spaces appear (_ _ _) with an asterisk above or below one of the spaces. If the asterisk is

above the space, the participant is to add 1 to the (imaginary) digit currently in that space, and if the asterisk is below the space, the participant is to subtract 1 from the digit currently in that space. For example, if on the second trial, an asterisk appears above the middle space, the new value is 565, if on the third trial the asterisk is below the third space, the new value is 564, and so on. After a random number of displays, a prompt appears, and the participant is to report the updated value.

Appendix D.1. Psychometric Network Analyses

We repeated the psychometric network analyses shown in Figure 1, adding Mental Counters. Results are shown in Figure A4. As in the main analysis, the partial correlation between AFQT and Multitasking Performance is nonsignificant. Furthermore, the partial correlations of both Gf and Placekeeping Ability with Multitasking Performance are significant, whereas the partial correlations of both Attention Control and Mental Counters with Multitasking Performance are both nonsignificant. In sum, as in the main analysis, Gf and Placekeeping Ability contributed to the AFQT and Multitasking Performance relationship, whereas Attention Control did not.

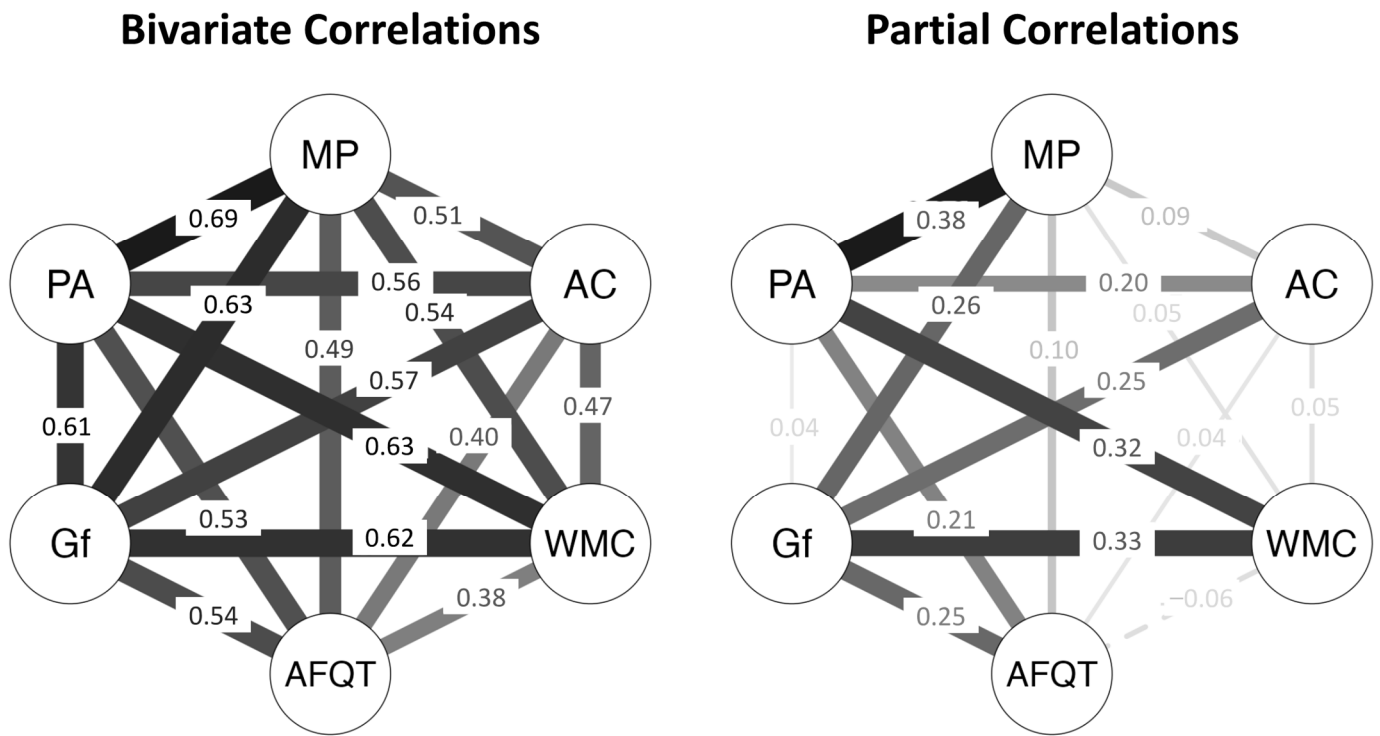


Figure A4. Psychometric network models with the Mental Counters measure of working memory capacity (WMC) as a predictor variable, along with composite variables from the original network model.

Appendix D.2. Hierarchical Regression Analyses

We performed hierarchical regression analyses with Mental Counters as a predictor variable. Results are presented in Table A5. Mental Counters improved prediction of Multitasking Performance over AFQT (Model 1, $\Delta R^2 = 0.142$, $\Delta F = 61.27$, $p < .001$, $\beta = 0.41$), and over AFQT and each of the other predictor variables: Gf (Model 2A, $\Delta R^2 = 0.028$, $\Delta F = 13.98$, $p < .001$, $\beta = 0.22$), Placekeeping Ability (Model 2B, $\Delta R^2 = 0.015$, $\Delta F = 8.23$, $p < .01$, $\beta = 0.16$), and Attention Control (Model 2C, $\Delta R^2 = 0.072$, $p < .001$, $\beta = 0.31$). However, Mental Counters did not improve the prediction of Multitasking Performance over AFQT and all three other predictor variables (Model 3, $\Delta R^2 = 0.001$, $\Delta F < 1$, $\beta = 0.05$).

Table A5. Results of Hierarchical Regression Analyses Predicting Multitasking Performance with Mental Counters as Predictor Variable.

	ΔR^2	ΔF	<i>df</i>	β [95% CI]	<i>t</i>
Model 1					
Step 1	0.241	85.17 ***	1, 268		
AFQT				0.49 [0.39, 0.60]	9.23 ***
Step 2	0.142	61.27 ***	1, 267		
MC				0.41 [0.31, 0.51]	7.83 ***
Model 2A					
Step 1	0.434	102.42 ***	2, 267		
AFQT				0.21 [0.10, 0.32]	3.83 ***
Gf				0.52 [0.41, 0.63]	9.54 ***
Step 2	0.028	13.98 ***	1, 266		
AFQT				0.20 [0.09, 0.30]	3.63 ***
Gf				0.40 [0.27, 0.52]	6.28 ***
MC				0.22 [0.10, 0.33]	3.74 ***
Model 2B					
Step 1	0.498	132.44 ***	2, 267		
AFQT				0.18 [0.08, 0.28]	3.45 ***
PA				0.60 [0.50, 0.70]	11.69 ***
Step 2	0.015	8.23 **	1, 266		
AFQT				0.16 [0.07, 0.26]	3.25 ***
PA				0.50 [0.39, 0.62]	8.44 ***
MC				0.16 [0.05, 0.27]	2.87 **
Model 2C					
Step 1	0.360	75.06 ***	2, 267		
AFQT				0.34 [0.23, 0.44]	6.33 ***
AC				0.38 [0.27, 0.48]	7.04 ***
Step 2	0.072	33.62 ***	1, 266		
AFQT				0.27 [0.16, 0.37]	5.10 ***
AC				0.26 [0.15, 0.37]	4.79 ***
MC				0.31 [0.21, 0.42]	5.80 ***
Model 3					
Step 1	0.557	83.15 ***	4, 265		
AFQT				0.08 [−0.02, 0.18]	1.54
Gf				0.28 [0.17, 0.40]	4.92 ***
PA				0.43 [0.32, 0.54]	7.59 ***
AC				0.08 [−0.02, 0.19]	1.54
Step 2	0.001	0.63	1, 264		
AFQT				0.08 [−0.02, 0.18]	1.58
Gf				0.27 [0.15, 0.39]	4.32 ***
PA				0.41 [0.30, 0.53]	6.78 ***
AC				0.08 [−0.03, 0.18]	1.49
MC				0.05 [−0.07, 0.16]	0.79

Note. *N* = 270. AFQT, Armed Forces Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability; AC, Attention Control; MC, Mental Counters. **, *p* < .01. ***, *p* < .001.

Appendix E

Here we report simultaneous regression analyses that evaluate the individual and combined contributions of the predictor variables to Multitasking Performance, to support an assessment of the amount of criterion variance explained as a function of a number of predictors.

Table A6 reports *R*² values and Figure A5 plots percent of variance explained (i.e., *R*² × 100). In the first set of models, we regressed Multitasking Performance onto each predictor variable individually. The largest effect size was for Placekeeping Ability (*R*² = 0.476). In the second set of models, we regressed Multitasking Performance onto all possible pairs of predictor variables. The largest effect size was for Gf and Placekeeping Ability (*R*² = 0.548). In the third set of models, we regressed Multitasking Performance onto all possible triads of predictor variables. The largest effect size (*R*² = 0.553) was for AFQT,

Gf, and Placekeeping Ability. Thus, within each set, the model with the highest variance explained includes Placekeeping Ability. In the fourth model, we regressed Multitasking Performance onto all four predictive variables ($R^2 = 0.557$). Altogether, then, the four predictor variables accounted for 56% of the variance in Multitasking Performance.

Table A6. Regression Analyses Predicting Multitasking Performance (1–4 Predictor Variables).

Model	R^2	F	df
1 predictor			
AFQT	0.241	85.17	1, 268
AC	0.264	96.11	1, 268
Gf	0.403	180.93	1, 268
PA	0.476	243.05	1, 268
2 predictors			
AFQT, AC	0.360	75.06	2, 267
AFQT, Gf	0.434	102.42	2, 267
Gf, AC	0.437	103.64	2, 267
AFQT, PA	0.498	132.44	2, 267
PA, AC	0.500	133.52	2, 267
Gf, PA	0.548	161.99	2, 267
3 predictors			
AFQT, Gf, AC	0.460	75.59	3, 266
AFQT, PA, AC	0.516	94.57	3, 266
Gf, PA, AC	0.533	109.50	3, 266
AFQT, Gf, PA	0.553	109.51	3, 266
4 predictors			
AFQT, Gf, PA, AC	0.557	83.15	4, 265

Note. $N = 270$. AFQT, Armed Forces Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability; AC, Attention Control. All F statistics are statistically significant, $p < .001$.

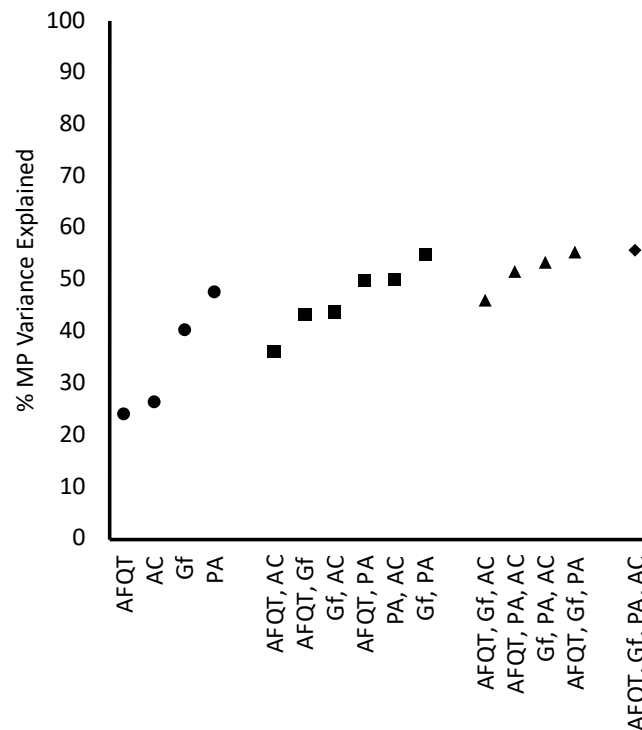


Figure A5. Percentage of variance ($R^2 \times 100$) in Multitasking Performance explained by predictor variables. MP, Multitasking Performance; AFQT, Armed Forces Qualification Test; Gf, Fluid Intelligence; PA, Placekeeping Ability; AC, Attention Control. Circles, 1-predictor models; squares, 2-predictor models; triangles, 3-predictor models; diamond, 4-predictor model.

Appendix F

Screenshots from Multitasking Paradigms.

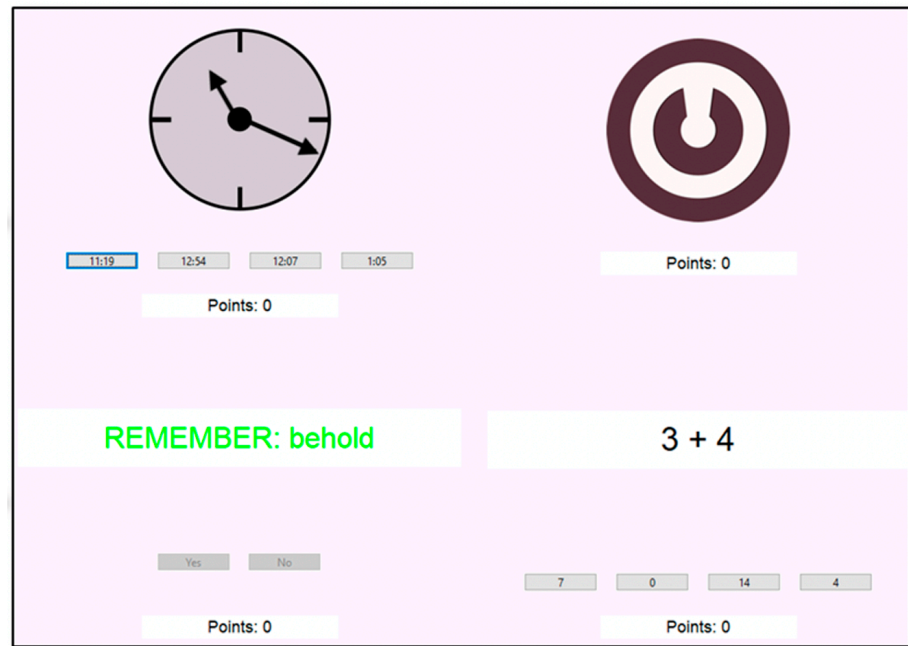


Figure A6. Foster Task.

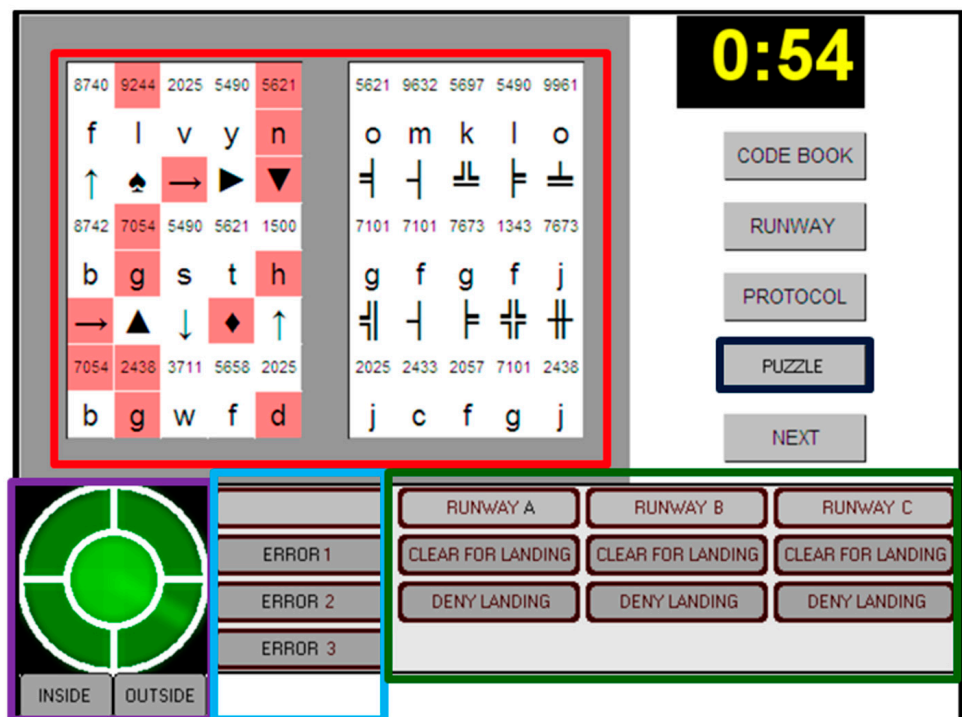


Figure A7. Control Tower.

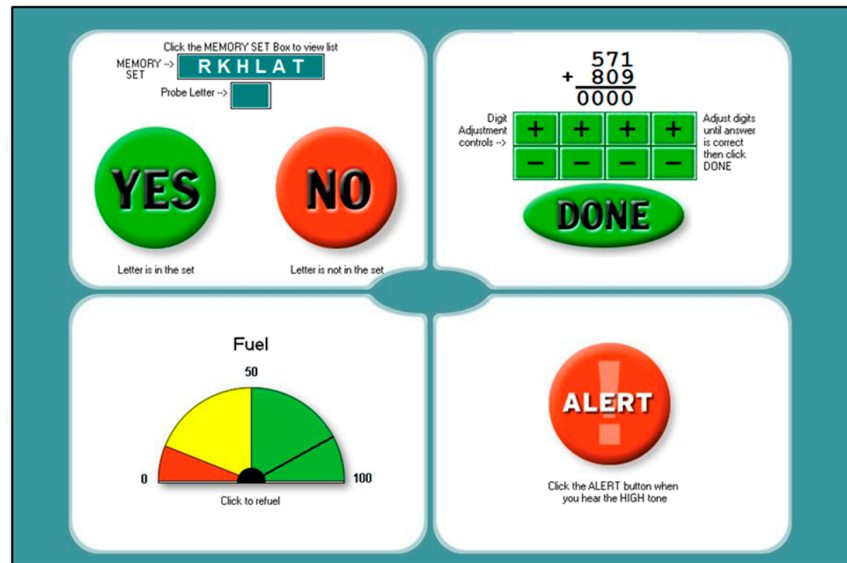
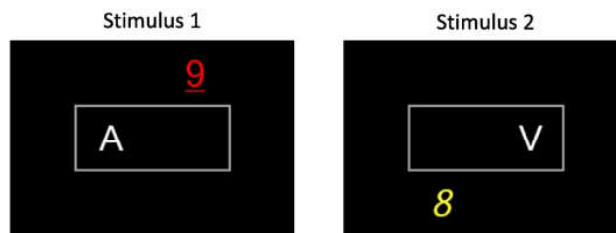


Figure A8. SynWin.

Screenshots from Placekeeping Tasks.

(a) Sample stimuli for placekeeping trials:



(b) Decision rule for each step, candidate choices for each rule, and correct choices for each rule given each stimulus in (a):

Step	Decision Rule	Candidate Responses	Correct Response for ...	
			Stimulus 1	Stimulus 2
U	a character is Underlined or in Italics	u i	u	i
N	letter is Near to or Far from start of alphabet	n f	n	f
R	a character is Red or Yellow	r y	r	y
A	a character is Above or Below the box	a b	a	b
V	letter is a Vowel or a Consonant	v c	v	c
E	digit is Even or Odd	e o	o	e
L	digit is Less than or More than 5	l m	m	m

(c) Sample stimulus and partial response for the typing task during interruptions:

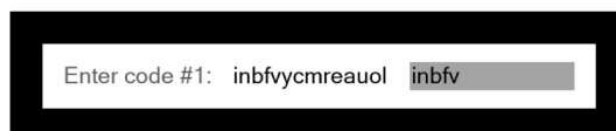
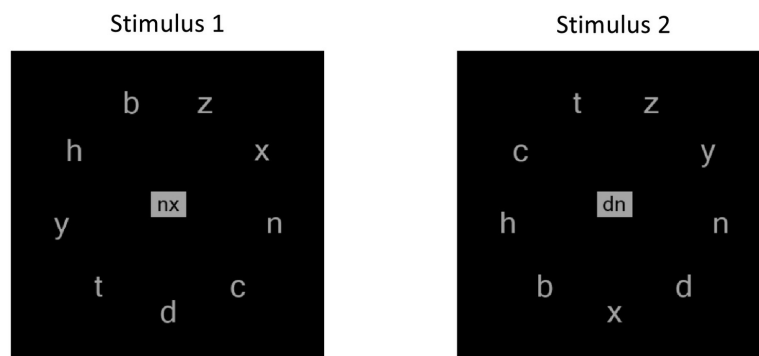


Figure A9. UNRAVEL task. In (a) the character 9 is displayed in red, and the character 8 in yellow (these characters are printed in color in the web version of this article). Stimuli for placekeeping trials, as in (a), are randomly generated for each placekeeping trial (Altmann et al. 2014). During interruptions, participants enter two separate “codes” like that in (c), each a random permutation of the 14 candidate responses identified in (b). Figure and description from Altmann and Hambrick (2022).

(a) Sample stimuli and partial responses for two consecutive placekeeping trials:



(b) Sample stimulus for the counting task during interruptions:

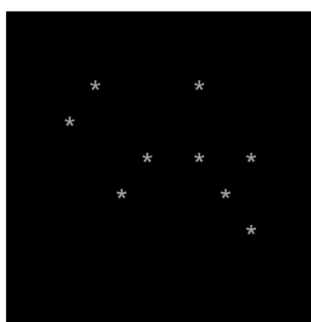


Figure A10. Letterwheel task. Stimuli for placekeeping trials, as in (a), are generated for each trial by randomly assigning the same set of letters to different positions around the wheel. The partial response for Stimulus 1, in the center of the wheel, is correct (with the final letter z) if the to-be-alphabetized letters are z, x, and n. The partial response for Stimulus 2 is correct (with the final letter y) if this stimulus is from the subsequent trial, reflecting a shift of one position clockwise for the alphabetizing operation. During interruptions, participants count sets of asterisks like that in (b). There are 7 to 9 asterisks per set, randomly distributed in the same region of the display as the letter wheel, and 5 sets per interruption. Figure and description from [Altmann and Hambrick \(2022\)](#).

Notes

- ¹ In SEM, a latent variable is represented as a circle (or oval) and an observed variable as a square (or rectangle). In psychometric network analysis, a node is represented as a circle, but this may not indicate that the variable is latent. The nodes in the psychometric network analyses we present here are not latent variables, but rather are composites of observed variables.
- ² Our battery included one test of working memory capacity (Mental Counters, administered in Session 2). Because we had only this one indicator of working memory capacity, we did not include it in our main analyses, and instead report a separate analysis including results of this test in Appendix D. We are grateful to Cyrus Foroughi, Ph.D., U.S. Naval Research Lab, for sending us this test and granting us permission to use it.
- ³ At present, the minimum AFQT percentile score (with a high school degree) is 30 for the Army, 35 for the Navy, 32 for the Marine Corps, 31 for the Air Force, and 36 for the Coast Guard (<https://www.military.com/join-armed-forces/asvab>, accessed on 3 December 2023).
- ⁴ We are extremely grateful to Randy Engle for granting us permission to use these attention control tasks in our study and to Jason Tsukahara for sending the tasks to us.

References

- Ackerman, Philip L., Margaret E. Beier, and Mary O. Boyle. 2005. Working memory and intelligence: The same or different constructs? *Psychological Bulletin* 131: 30. [[CrossRef](#)] [[PubMed](#)]
- Alluisi, Earl A. 1967. Methodology in the use of synthetic tasks to assess complex performance. *Human Factors* 9: 375–84. [[CrossRef](#)] [[PubMed](#)]

- Altmann, Erik M., and David Z. Hambrick. 2022. Task independence of placekeeping as a cognitive control construct: Evidence from individual differences and experimental effects. *Cognition* 229: 105229. [[CrossRef](#)] [[PubMed](#)]
- Altmann, Erik M., J. Gregory Trafton, and David Z. Hambrick. 2014. Momentary interruptions can derail the train of thought. *Journal of Experimental Psychology: General* 143: 215–26. [[CrossRef](#)] [[PubMed](#)]
- Banich, Marie T. 2009. Executive function: The search for an integrated account. *Current Directions in Psychological Science* 18: 89–94. [[CrossRef](#)]
- Borsboom, Denny, Marie K. Deserno, Mijke Rhemtulla, Sacha Epskamp, Eiko I. Fried, Richard J. McNally, Donald J. Robinaugh, Marco Perugini, Jonas Dalege, Giulio Costantini, and et al. 2021. Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers* 1: 58. [[CrossRef](#)]
- Burgoyne, Alexander P., and Randall W. Engle. 2020. Attention control: A cornerstone of higher-order cognition. *Current Directions in Psychological Science* 29: 624–30. [[CrossRef](#)]
- Burgoyne, Alexander P., Cody A. Mashburn, and Randall W. Engle. 2021a. Reducing adverse impact in high-stakes testing. *Intelligence* 87: 101561. [[CrossRef](#)]
- Burgoyne, Alexander P., David Z. Hambrick, and Erik M. Altmann. 2019. Placekeeping ability as a component of fluid intelligence: Not just working memory capacity. *American Journal of Psychology* 132: 439–49. [[CrossRef](#)]
- Burgoyne, Alexander P., David Z. Hambrick, and Erik M. Altmann. 2021b. Incremental validity of placekeeping as a predictor of multitasking. *Psychological Research* 85: 1515–28. [[CrossRef](#)]
- Burgoyne, Alexander P., Jason S. Tsukahara, Cody A. Mashburn, Richard Pak, and Randall W. Engle. 2023. Nature and measurement of attention control. *Journal of Experimental Psychology: General* 152: 2369–402. [[CrossRef](#)] [[PubMed](#)]
- Cattell, Raymond B. 1943. The measurement of adult intelligence. *Psychological Bulletin* 40: 153–93. [[CrossRef](#)]
- Cattell, Raymond B. 1987. *Abilities: Their Structure, Action, and Growth*. Boston: Houghton Mifflin. First published 1971.
- Colom, Roberto, Agustín Martínez-Molina, Pei C. Shih, and José Santacreu. 2010. Intelligence, working memory, and multitasking performance. *Intelligence* 38: 543–51. [[CrossRef](#)]
- Conway, Andrew R. A., Nelson Cowan, Michael F. Bunting, David J. Theriault, and Scott R. Minkoff. 2002. A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence* 30: 163–83. [[CrossRef](#)]
- Deary, Ian. 2000. *Looking Down on Human Intelligence: From Psychometrics to the Brain*. Oxford: OUP Oxford, vol. 36.
- Draheim, Christopher W., Jason S. Tsukahara, Jessie D. Martin, Cody A. Mashburn, and Randall W. Engle. 2021. A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General* 150: 242–75. [[CrossRef](#)] [[PubMed](#)]
- Ekstrom, Ruth B., John W. French, Harry H. Harman, and Diran Dermen. 1976. *Kit of Factor-Reference Cognitive Tests*. Princeton: Educational Testing Service.
- Elsmore, Timothy F. 1994. SYNWORK1: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instruments, & Computers* 26: 421–26. [[CrossRef](#)]
- Hambrick, David Z., Alexander B. Burgoyne, and Frederick L. Oswald. 2023. The validity of general cognitive ability predicting job-specific performance is stable across different levels of job experience. *Journal of Applied Psychology*. [[CrossRef](#)]
- Hambrick, David Z., Tara A. Rench, Elizabeth M. Poposki, Emily S. Darowski, Donna Roland, Ronald M. Bearden, Frederick L. Oswald, and Randy Brou. 2011. The relationship between the ASVAB and multitasking in Navy sailors: A process-specific approach. *Military Psychology* 23: 365–80. [[CrossRef](#)]
- Hasher, Lynn, Cynthia Lustig, and Rose T. Zacks. 2007. Inhibitory mechanisms and the control of attention. In *Models of Working Memory: Mechanisms of Active Maintenance and Executive Function*. Edited by Akira Miyake and Priti Shah. Cambridge: Cambridge University Press, pp. 227–49.
- Held, Janet D., and John H. Wolfe. 1997. Validities of unit-weighted composites of the ASVAB and the ECAT battery. *Military Psychology* 9: 77–84. [[CrossRef](#)]
- Held, Janet D., Sarah A. Hezlett, Jeff W. Johnson, Rodney A. McCloy, Fritz Drasgow, and Eduardo Salas. 2014. *Introductory Guide for Conducting ASVAB Validation/Standards Studies in the US Navy*. Millington: Navy Personnel Research Studies and Technology.
- Hunt, Earl. 2010. *Human Intelligence*. Cambridge: Cambridge University Press. [[CrossRef](#)]
- Hunter, John E., and Frank L. Schmidt. 1996. Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law* 2: 447–72. [[CrossRef](#)]
- Jastrzębski, Jan, Michal Ociepka, and Adam Chuderski. 2023. Graph Mapping: A novel and simple test to validly assess fluid reasoning. *Behavior Research Methods* 55: 448–60. [[CrossRef](#)] [[PubMed](#)]
- Kane, Michael J., David Z. Hambrick, Stephen W. Tuholski, Oliver Wilhelm, Tabitha W. Payne, and Randall W. Engle. 2004. The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General* 133: 189–217. [[CrossRef](#)] [[PubMed](#)]
- Koch, Iring, Edita Poljac, Hermann Müller, and Andrea Kiesel. 2018. Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin* 144: 557–83. [[CrossRef](#)] [[PubMed](#)]
- Kovacs, Kristof, and Andrew R. A. Conway. 2016. Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry* 27: 151–77. [[CrossRef](#)]

- Kyllonen, Patrick C. 1993. Aptitude testing inspired by information processing: A test of the four-sources model. *The Journal of General Psychology* 120: 375–405. [CrossRef]
- Kyllonen, Patrick C., and Raymond E. Christal. 1990. Reasoning ability is (little more than) working-memory capacity?! *Intelligence* 14: 389–433. [CrossRef]
- Larson, Gerald E., and David L. Alderton. 1990. Reaction time variability and intelligence: A “worst performance” analysis of individual differences. *Intelligence* 14: 309–25. [CrossRef]
- Larson, Gerald E., and John H. Wolfe. 1995. Validity results for *g* from an expanded test base. *Intelligence* 20: 15–25. [CrossRef]
- Lee, Michael D., and Eric J. Wagenmakers. 2014. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge: Cambridge University Press.
- Martin, Jessie D., Cody A. Mashburn, and Randall W. Engle. 2020. Improving the validity of the Armed Services Vocational Aptitude Battery with measures of attention control. *Journal of Applied Research in Memory and Cognition* 9: 323–35. [CrossRef]
- Mayberry, Paul W. 1990. *Validation of ASVAB against Infantry Job Performance*. Alexandria: Center for Naval Analyses, Force Structure and Acquisition Division.
- McGrew, Kevin S. 2023. Carroll’s three-stratum (3S) cognitive ability theory at 30 years: Impact, 3S-CHC theory clarification, structural replication, and cognitive–achievement psychometric network analysis extension. *Journal of Intelligence* 11: 32. [CrossRef] [PubMed]
- Morgan, Benjamin B., Jr., and Earl A. Alluisi. 1972. Synthetic work: Methodology for assessment of human performance. *Perceptual and Motor Skills* 35: 835–45. [CrossRef]
- National Center for Education Statistics. 2019. *Digest of Education Statistics*. Available online: https://nces.ed.gov/programs/digest/d19/tables/dt19_226.40.asp (accessed on 29 June 2023).
- National Center for O*NET Development. 2023. O*NET OnLine. Available online: <https://www.onetonline.org/link/summary/53-2021.00> (accessed on 4 February 2023).
- Raven, John, John C. Raven, and John H. Court. 1998. Manual for Raven’s progressive matrices and vocabulary scales: Section 4. In *The Advanced Progressive Matrices*. San Antonio: Harcourt Assessment.
- Redick, Thomas S., Zachary Shipstead, Matthew E. Meier, Janelle J. Montroy, Kenny L. Hicks, Nash Unsworth, Michael J. Kane, David Z. Hambrick, and Randall W. Engle. 2016. Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General* 145: 1473–92. [CrossRef] [PubMed]
- Ree, Malcolm J., and James A. Earles. 1992. Intelligence is the best predictor of job performance. *Current Directions in Psychological Science* 1: 86–89. [CrossRef]
- Ree, Malcolm J., and Thomas R. Carretta. 2022. Thirty years of research on general and specific abilities: Still not much more than *g*. *Intelligence* 91: 101617. [CrossRef]
- Ree, Malcolm J., and Toni G. Wegner. 1990. Correcting differences in answer sheets for the 1980 Armed Services Vocational Aptitude Battery reference population. *Military Psychology* 2: 157–69. [CrossRef]
- Ree, Malcolm J., James A. Earles, and Mark S. Teachout. 1994. Predicting job performance: Not much more than *g*. *Journal of Applied Psychology* 79: 518–24. [CrossRef]
- Reeve, Charlie L., and Silvia Bonaccio. 2011. On the myth and the reality of the temporal validity degradation of general mental ability test scores. *Intelligence* 39: 255–72. [CrossRef]
- Roberts, Richard D., Ginger N. Goff, Fadi Anjoul, Patrick C. Kyllonen, Gerry Pallier, and Lazar Stankov. 2000. The Armed Services Vocational Aptitude Battery (ASVAB): Little more than acculturated learning (Gc)!? *Learning and Individual Differences* 12: 81–103. [CrossRef]
- Schmank, Christopher J., Sara A. Goring, Kristof Kovacs, and Andrew R. A. Conway. 2019. Psychometric network analysis of the Hungarian WAIS. *Journal of Intelligence* 7: 21. [CrossRef]
- Schmidt, Frank L., and John Hunter. 2004. General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology* 86: 162–73. [CrossRef] [PubMed]
- Schmidt, Frank L., John E. Hunter, Alice N. Outerbridge, and Stephen Goff. 1988. Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology* 73: 46–57. [CrossRef]
- Schmidt, Frank L., John E. Hunter, Robert C. McKenzie, and Tressie W. Muldrow. 1979. Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology* 64: 609–26. [CrossRef]
- Thurstone, Louis L. 1938. *Primary Mental Abilities*. Chicago: University of Chicago Press.
- Wall, Janet E. 2017. *ASVAB Basic Training for the AFQT*. New York: McGraw-Hill.
- Welsh, John R., Susan K. Kucinkas, and Linda T. Curran. 1990. Armed Services Vocational Battery (ASVAB): Integrative review of validity studies. Available online: <https://apps.dtic.mil/sti/pdfs/ADA225074.pdf> (accessed on 4 December 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.