International Journal of Selection and Assessment

WILEY

International Journal of SELECTION AND ASSESSMENT

RESEARCH ARTICLE

Attention Control Measures Improve the Prediction of Performance in Navy Trainees

Alexander P. Burgoyne[1,2] | Cody A. Mashburn[1] | Jason S. Tsukahara[1] | Richard Pak[3] | Joseph T. Coyne[4] | Cyrus Foroughi[4] | Ciara Sibley[4] | Sabrina M. Drollinger[5] | Randall W. Engle[1]

[1]School of Psychology, Georgia Institute of Technology, Atlanta, Georgia, USA | [2]Human Resources Research Organization (HumRRO), Alexandria, Virginia, USA | [3]College of Behavioral, Social and Health Sciences, Clemson University, Clemson, South Carolina, USA | [4]Information Technology Division, Naval Research Laboratory, Washington, DC, USA | [5]Naval Aerospace Medical Institute, Pensacola, Florida, USA

Correspondence: Alexander P. Burgoyne (burgoyn4@gmail.com)

Received: 22 January 2024 | **Revised:** 22 October 2024 | **Accepted:** 30 October 2024

Funding: This work was supported by a grant from the Naval Research Laboratory N00173-20-2-C003 to Randall W. Engle and a grant from the Office of Naval Research N00014-23-1-2768 to Randall W. Engle and Alexander P. Burgoyne.

Keywords: air traffic control | Armed Forces Qualificataion Test | attention control | personnel selection

ABSTRACT

Military selection tests leave room for improvement when predicting work-relevant outcomes. We tested whether measures of attention control, working memory capacity, and fluid intelligence improved the prediction of training success above and beyond composite scores used by the U.S. Military. For student air traffic controllers, commonality analyses revealed that attention control explained 9.1% ($R = .30$) of the unique variance in academic performance, whereas the Armed Forces Qualification Test explained 5.2% ($r = .23$) of the unique variance. For student naval aviators, incremental validity estimates were small and nonsignificant. For student naval flight officers, commonality analyses revealed that attention control measures explained 11.8% ($R = .34$) of the unique variance in aviation preflight indoctrination training performance and 4.3% ($R = .21$) of the unique variance in flight performance. Although these point estimates are based on relatively small samples, they provide preliminary evidence that attention control measures might improve training outcome classification accuracy in real-world samples of military personnel.

1 | Introduction

When performing any cognitive task, one must contend with distractions, interference, impulses, and mind wandering. *Attention control* is what allows us to focus on goal-relevant information while ignoring or suppressing these distracting thoughts and events. Because distractions and interference are ubiquitous in everyday life, individual differences in the ability to control attention have far reaching consequences for real-world cognitive performance (Dempster 1991; Engle 2002, 2018; Mashburn, Burgoyne, and Engle 2023).

Individual differences in attention control explain considerable variance in academic achievement (Ahmed et al. 2019; Blankenship et al. 2019; McVay and Kane 2012), job performance (Nelson 2003; Guo et al. 2020; Bosco, Allen, and Singh 2015), emotion regulation (Engen and Anderson 2018; Garrison and Schmeichel 2020; Moran 2016), rationality (Burgoyne, Mashburn, and Engle 2021), sensory discrimination ability (Tsukahara et al. 2020), and complex problem solving (Dempster 1991; Draheim et al. 2021; Martin, Mashburn, and Engle 2020; Salthouse, Atkinson, and Berish 2003). For reviews of the expansive role of attention control in complex behavior, see Mashburn, Burgoyne, and Engle (2023) and Draheim et al. (2022). Based on these results, it has been argued that attention control is a domain-general cognitive ability that plays a role in most controlled mental operations, particularly under conditions in which interference leads to the retrieval of response

Summary

- We tested whether measures of attention control improve the prediction of Navrfy training performance.
- For student air traffic controllers, commonality analyses revealed that attention control explained 9.1% of the unique variance in academic performance, compared to 5.2% by the Armed Forces Qualification Test.
- For student naval flight officers, commonality analyses revealed that attention control explained 11.8% of the unique variance in aviation preflight indoctrination training performance and 4.3% of the unique variance in flight performance.

tendencies that conflict with current task goals (Burgoyne and Engle 2020; Engle 2002; Engle 2018). The view that attention control is central to complex cognition can be traced back to Titchener (1908), who stated that "the doctrine of attention is the nerve of the whole psychological system" (p. 173).

Extending this argument, Burgoyne et al. (2022) recently tested whether attention control explains the *positive manifold*—the positive correlations observed among broad cognitive abilities. The data set included measures of attention control, working memory capacity, fluid intelligence, and sensory discrimination ability (see Tsukahara et al. 2020). Using a two-step modeling approach, Burgoyne et al. (2022) found that attention control had the highest loading on the *g*-factor, the latent variable which provides a statistical explanation for the positive manifold. Next, the researchers specified attention control as a predictor of the other cognitive ability factors, and tested whether the correlations among their residuals were reduced to a meaningful degree. Although attention control did not fully eliminate the residual correlations, it did reduce them considerably. Burgoyne et al. (2022) interpreted this as evidence that attention control is a common element that contributes to individual differences in a variety of complex cognitive functions. That is, attention control can be seen as a bottleneck that constrains performance across cognitive domains; if one generally struggles to control their attention, performance on specific tests of reasoning, memory, perceptual speed, and so on will be negatively affected as a result.

The purpose of this paper is to investigate whether measures of attention control predict individual differences in training performance in the U.S. military. Presumably, "zoning out" or not paying attention will have negative consequences for knowledge and skill acquisition (Unsworth et al. 2012). Indeed, in a sample of Air Force trainees, Woltz (1988) found that individual differences in attention control predicted declarative rule acquisition and proceduralization in a complex task meant to simulate work performed at a control panel.

Attention control also plays a role in academic achievement. For example, Unsworth et al. (2012) found that individual differences in attention control predicted everyday attentional failures such as mind wandering during class, which in turn predicted performance on the SAT. McVay and Kane (2012) found that the ability to control attention was significantly correlated with the number of task-unrelated thoughts students experienced while reading, and furthermore, predicted passage comprehension.

Attention control also plays a role in motor skill acquisition. As a case in point, Burgoyne, Harris and Hambrick (2019) challenged novice pianists to learn a new piece of music given 12 min of practice. They found that participants with greater working memory capacity (i.e., a proxy measure for attention control; see Engle et al. 1999; Conway et al. 2002) performed significantly better in the skill acquisition task than individuals with lower working memory scores. Attention control contributes substantially to individual differences in working memory performance, and largely explains its predictive validity and relation to other cognitive constructs, such as fluid intelligence (Engle et al. 1999; Conway et al. 2002; Draheim et al. 2021). Thus, attention control appears to be important for both academic skill acquisition and procedural skill acquisition, both of which might be important to military training programs.

Our argument is that cognitive science and psychometrics can and should be used to guide applied psychology. Cognitive science can identify theoretical constructs that are important for particular jobs and isolate the key elements or cognitive processes captured by tests that are responsible for their relationship to job performance. Psychometrics can be used to improve reliability and subsequent validity. Recent developments in cognitive psychology (e.g., solutions to the "reliability paradox"; Hedge, Powell, and Sumner 2018, p. 1166) have yet to be adopted in the applied sector, delaying progress. Unfortunately, this observation is not new; 30 years ago, Landy, Shankster and Kohler (1994) stated: "It is with some embarrassment, then, that we must recognize that little progress is apparent in the conception and understanding of cognitive abilities by I/O psychologists after more than 100 years of mental testing" (pp. 267–268). Although progress has been made since then, strengthening the link between cognitive psychology and industrial/organizational (I/O) psychology will accelerate discoveries and foster cross-pollination with practical utility.

1.1 | Applications to Military Personnel Selection

The U.S. military has invested millions of dollars and decades of research developing assessments to select and classify personnel. Perhaps the most well-known of these assessments is the Armed Services Vocational Aptitude Battery (ASVAB), a series of cognitive ability tests administered to more than one million people each year, including every enlisted applicant (ASVAB Enlistment Testing Program 2020a). The ASVAB and the Armed Forces Qualification Test (AFQT), which comprises the math and verbal subtests, are critical to the military's goal of selecting, developing, and retaining the right people for the job (Dempsey 2015).

Scores on military selection tests predict training success, but they leave room for improvement. For example, a meta-analysis of data from Valentine (1977) indicated that the AFQT correlated $r = .34$ ($R^2 = 12\%$) with final school grades in a sample of 43,985 Air Force trainees (Welsh, Kucinkas, and Curran 1990). In an even larger sample of 2,476,608, Wegner and Ree (1986) found a meta-analytic average correlation of $r = .42$ ($R^2 = 18\%$) between AFQT scores and final class grades.

The results are similar when considering the relationship between AFQT scores and military job performance. For example, using archival data from the Joint-Service Job Performance Measurement/Enlistment Standards (JPM) Project ($N = 10{,}088$), Hambrick, Burgoyne and Oswald (2024) found that the meta-analytic average correlation between AFQT scores and hands-on job performance test scores was $r = 0.24$, or $r = 0.39$ after correcting for range restriction and criterion unreliability. For context, the measure of hands-on job performance was "based on the percentage of MOS-specific task steps that the soldier was observed to perform successfully (for further information on the HOPT measures, see Wigdor and Green 1991; Wise 1994)" (Hambrick, Burgoyne, and Oswald 2024, p. 440). Time in service predicted job performance to a comparable degree (meta-analytic $r = 0.25$, or $r = 0.27$ after correcting for criterion unreliability). Multiple regression analyses revealed that, on average, AFQT scores and time in service explained less than one-fifth of the variance in hands-on job performance test scores, and their interaction was negligible.

Although no selection test can reasonably be expected to explain all the variance in work-relevant performance, even small improvements in predictive validity can have large consequences for trainee retention and organizational effectiveness (Held, Carretta, and Rumsey 2014; Schmidt, Dunn, and Hunter 1995).

A second issue when considering the merits of a selection test is its potential for negative societal consequences, such as adverse impact. *Adverse impact* refers to the disproportionate selection or promotion of one protected class of individuals over another (Zedeck 2011). As the mean difference between groups' scores on a selection test increases, so does the likelihood of adverse impact. In the United States, adverse impact is a legal issue. If a selection test has adverse impact, the hiring organization must demonstrate that the test is valid, relevant to the job, and that alternatives have been explored (Uniform Guidelines On Employee Selection Procedures 1978). That said, the U.S. military is given some leeway when it comes to adverse impact; they are not technically bound by the same rules as other organizations in the United States (Kamarck 2019; Westergard 2019).

Nevertheless, the ASVAB results in adverse impact. Specifically, the qualification rate for Black applicants is less than 80% of the qualification rate for White applicants because of group differences in test performance (ASVAB Enlistment Testing Program 2020b; Wise et al. 1992). Once applicants have the minimum score to qualify for the Navy, they then qualify for different ratings (i.e., jobs) within the Navy based on composite scores from the ASVAB. We have access to ASVAB scores for all Navy enlisted personnel (over 329,000) who took the test from 2014 to 2022. These data are for those currently in the Navy—not all applicants. The air traffic control rating is of particular interest since it is one of the populations that participated in the present study. An analysis of all enlisted personnel from every rating has 31% of females qualifying for the air traffic control rating whereas 52% of men did. The percentage of African American sailors who qualify for the air traffic control rating is 25% compared to 58% of white applicants.

1.2 | The ASVAB and Crystallized Intelligence

Efforts to improve the ASVAB and AFQT have identified one issue that appears relevant to both improving predictive validity and reducing adverse impact. Current selection tests are heavily weighted towards *crystallized intelligence* (i.e., acquired knowledge). Roberts et al. (2000) conducted a factor analysis on the ASVAB and other cognitive ability tests and found that the ASVAB disproportionately emphasized acculturated learning. Tests of acquired knowledge are sensitive to differential access to quality education or specialized knowledge, socioeconomic factors, and interests (Bosco, Allen, and Singh 2015; Held, Carretta, and Rumsey 2014; Outtz and Newman 2011). Perhaps as a result of systemic inequalities in the United States, crystallized intelligence tests have larger group differences in performance than tests of other cognitive abilities. For instance, Outtz and Newman (2011) found that the subtests of the ASVAB with the largest differences between groups were those that measured verbal abilities and academic/technical knowledge (e.g., general science, auto and shop information). In a review, Hough, Oswald and Ployhart (2001) reported that tests of crystallized intelligence, and in particular, tests of science achievement and quantitative ability, had the largest differences between majority/minority ethnic groups. The evidence suggests that emphasizing acquired knowledge has exacerbated group differences in selection test performance and subsequent placement and promotion.

Nevertheless, knowledge is relevant to many occupations. We are not suggesting that crystallized intelligence tests be abandoned by the military, as they likely capture job-relevant knowledge in place before training begins. That said, other cognitive abilities may be increasingly important to military training success and job performance and may also generate smaller group differences. For example, many military occupations now require learning sophisticated and novel material, complex problem solving, and logical thinking (Held, Carretta, and Rumsey 2014). Beyond measuring what candidates already know, organizations should consider measuring their ability to learn, figure things out, or attend to task-relevant information while filtering out distractions and interference.

1.3 | Augmenting the ASVAB

The past few decades have seen a surge of interest in augmenting military selection tests with non-crystallized measures, such as tests of spatial abilities, problem solving, working memory, and attention control. For example, Held, Carretta and Rumsey (2014) tested whether two ASVAB subtests, "Assembling Objects" and "Coding Speed," added incremental validity to the prediction of Navy sailors' final school grades. Held, Carretta and Rumsey (2014) found that Coding Speed and Assembling Objects each accounted for a validity increment of around $r = 0.02$ compared to the current ASVAB composite scores with the highest validity for each occupation. To examine the practical effect of these seemingly small improvements, they conducted a hypothetical cost–benefit analysis for air traffic controllers in training. They found that by increasing predictive validity by just $r = 0.02$, 15 fewer of every 1000 sailors would fail training. At a cost of $100,000 per enlistee, the Navy would save $1.5 million on air traffic controllers alone. Furthermore, the Coding Speed and Assembling Objects subtests yielded smaller group differences than the technical knowledge subtests currently included in the ASVAB.

As a precursor to the present study, we investigated whether tests of attention control and fluid intelligence improved the prediction of multitasking performance above and beyond the ASVAB (Martin, Mashburn, and Engle 2020). We used a multitasking paradigm as a proxy for real-world work because many military occupations require performing multiple tasks simultaneously (or concurrently) (Burgoyne, Mashburn, and Engle 2021). In a sample of 171 young adult civilians recruited from Georgia Tech and the surrounding Atlanta community, we found that the ASVAB accounted for 77% of the variance in multitasking performance on its own. Note that in these analyses, there was no correction performed for range restriction on the ASVAB or the other ability measures—the sample consisted of civilians, not personnel who had been selected based on their test performance. Next, we estimated relationships between the ASVAB, attention control, and fluid intelligence at the latent level. There was considerable overlap; ASVAB performance was highly correlated with fluid intelligence ($r = 0.88$) and attention control (rs ranged from 0.60 to 0.71). Nevertheless, structural equation modeling revealed that attention control and fluid intelligence largely explained the ASVAB's predictive validity (Figure 1). Specifically, in the full model, attention control and fluid intelligence significantly predicted multitasking (βs = 0.45 and 0.47), but the ASVAB did not ($\beta = 0.16$, *ns*).

2 | Method

There are critical gaps in the literature that we attempted to address in the present study. For instance, only a handful of incremental validity studies have considered multiple cognitive ability constructs simultaneously to examine their relative

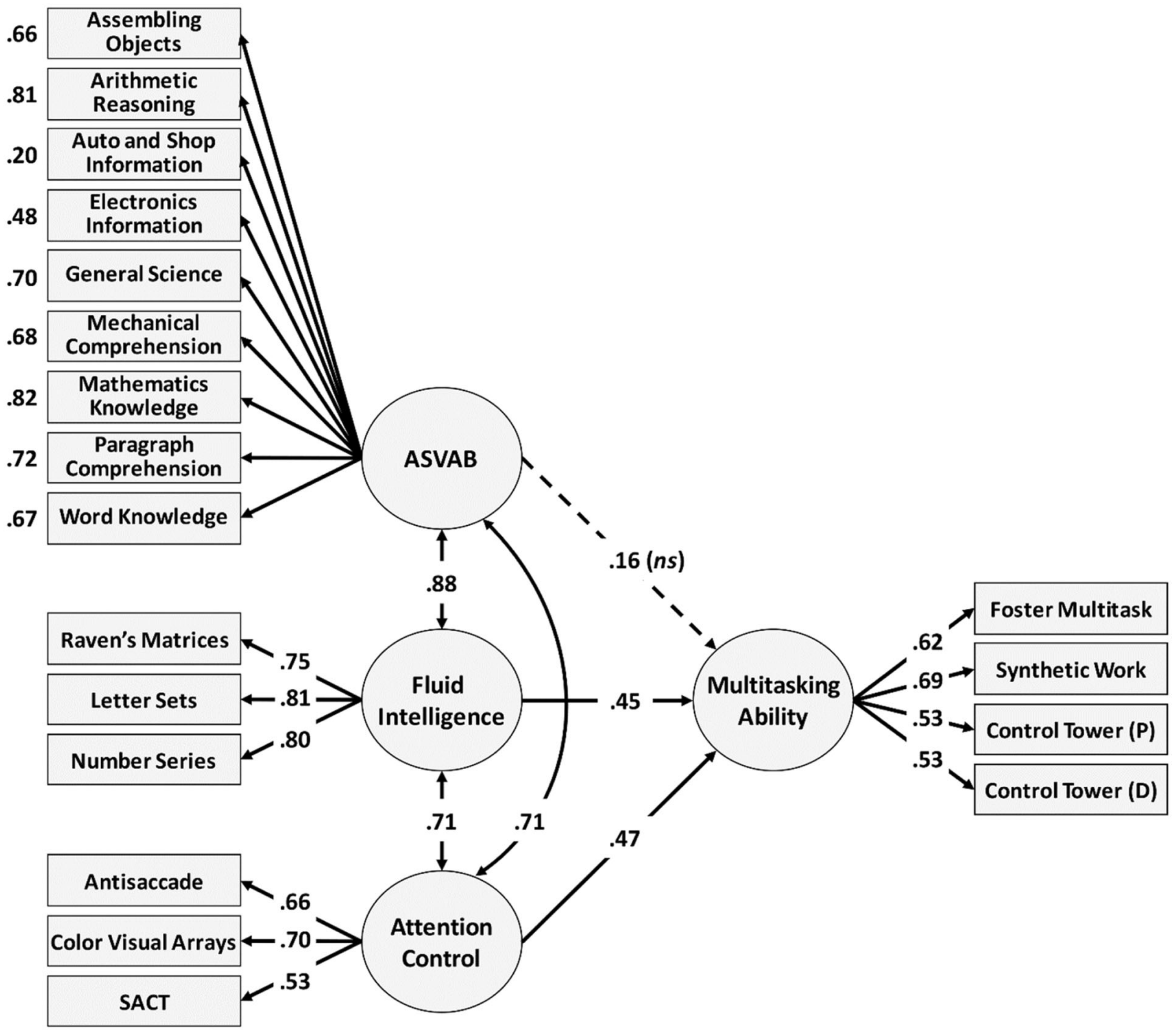

FIGURE 1 | Structural equation model with ASVAB scores, fluid intelligence, and attention control specified as correlated predictors of multitasking ability. Adapted from Martin, Mashburn, and Engle (2020). *Note:* Factor loadings for the ASVAB indicators are presented to the left side for visual clarity.

contributions, or their shared and unique contributions, to work-relevant performance. To what extent do fluid intelligence, working memory capacity, and attention control capture *unique* variance in work-relevant outcomes above and beyond one another and current selection tests? From a theoretical standpoint, this work can identify the cognitive constructs that account for the relationship between ability tests and outcome measures. From a practical perspective, it can shed light on which constructs (or tests) ought to be prioritized given limited testing time to maximize predictive validity.

Additionally, there is a pressing need to improve personnel selection for some of the Navy Enlisted Classifications (i.e., occupations). Specifically, the training success rate for air traffic controllers is surprisingly low; from 2014 to 2022, of the 3447 air traffic control students, only 41% passed without having an academic setback. Because air traffic control, piloting, and the work of naval flight officers requires attending to multiple streams of information concurrently, we reasoned that tests of attention control might improve the prediction of training success. Furthermore, given the relationship between attention control and learning rates for knowledge and skills, we predicted that attention control would explain variance in academic components of training (e.g., performance in the classroom) and procedural components of training (e.g., flying a plane).

To address these issues, we report the results of an ongoing collaboration between the Attention & Working Memory Lab at Georgia Tech, the Naval Research Lab, and the Naval Aerospace Medical Institute. We were interested in two questions: (1) whether tests of attention control (or tests of fluid intelligence and working memory capacity) enhance the prediction of military-relevant training performance above and beyond current selection tests; and (2) whether these predictors capture unique or shared variance in outcome measures. To this end, we administered a series of cognitive ability tests to Sailors and Marines in training. As they made their way through training, we obtained measures of training success. For the air traffic controllers, we collected data on academic setbacks and attrition, and for the student naval aviators and student naval flight officers, we collected data on their performance in the preliminary aviation preflight indoctrination (API) academic program and subsequent primary flight training and academic training performance. We also obtained their official scores on selection tests, including the ASVAB for enlisted personnel and composite scores based on subtests of the Aviation Selection Test Battery (ASTB) which are used to make selection decisions for aviators in training, such as the Academic Qualifications Rating (AQR), Pilot Flight Aptitude Rating (PFAR), and Flight Officer Flight Aptitude Rating (FOFAR). We used hierarchical regression analyses to estimate the incremental validity of the cognitive ability measures and commonality analyses to identify the unique and shared proportion of variance explained by each predictor (for more detail on commonality analyses, see Nimon 2010; Nimon and Oswald 2013).

2.1 | Participants

Our sample consisted of 490 U.S. Navy Sailors and Marines who had been selected for training at the Naval Aerospace Medical Institute (NAMI) in Pensacola, Florida for occupations in aviation and air traffic control. All participants provided informed consent.

2.2 | Sample Size Justification and Power Analysis

Our initial goal for data collection was to recruit as many trainees as we could from the Naval Aerospace Medical Institute. However, the COVID-19 pandemic created a major challenge and forced us to stop data collection in 2020 in the interest of public safety. The sample included in the present manuscript represents all participants who completed the study before the pandemic. Our effective sample consists of 119 air traffic controllers in training, 293 student naval aviators, and 78 student naval flight officers.

We performed a series of power analyses for the three samples in the present study using G*Power (Faul et al. 2009). Hambrick, Burgoyne, and Oswald (2024) report a meta-analytic average correlation of $r = 0.24$ between observed AFQT scores and hands-on job performance test scores based on data from the Joint-Service Job Performance Measurement/Enlistment Standards (JPM) Project. We use $r = 0.24$ as our first benchmark and report our power to detect correlations of this magnitude given the size of our three samples. For the 119 air traffic controllers in training, we have 75% power (two-tailed, alpha = 0.05); for the 293 student naval aviators, we have 99% power; for the 78 student naval flight officers, we have 57% power to detect a correlation of $r = 0.24$.

Using G*Power, we also tested our power to detect incremental validity in hierarchical regression models with specifications that matched the analyses reported later in the present manuscript (e.g., military selection measures entered in the first step; attention control measures entered in the second step). We used a baseline R^2 of 5.76% in the first step of these models based on the correlation between AFQT and job performance ($r = .24$) reported by Hambrick et al. (2011). For the 119 air traffic controllers in training, we have 23% power to detect an incremental validity of 2%, 55% power to detect an incremental validity of 5%, and 89% power to detect an incremental validity of 10%. For the 293 student naval aviators, we have 54% power to detect an incremental validity of 2%, 94% power to detect an incremental validity of 5%, and 99% power to detect an incremental validity of 10%. For the 78 student naval flight officers, we have 16% power to detect an incremental validity of 2%, 37% power to detect an incremental validity of 5%, and 70% power to detect an incremental validity of 10%.

2.3 | Procedure

After obtaining informed consent, participants completed computerized tests of cognitive ability in groups over the course of a single session. They were told that the tests they were completing were being considered as "special tests" to improve the prediction of performance among trainees, and therefore that it was important that they tried their best. The tasks were administered in a fixed order within each sample to avoid participant × order interactions (e.g., Hambrick et al. 2023).

2.4 | Demographics

Participants completed a demographic questionnaire which included items on age, gender, race/ethnicity, educational attainment, vision, handedness, and videogame experience.

2.5 | Attention Control

The attention control measures used in the present study required subjects to maintain focus on task-relevant information while ignoring or suppressing the influence of distractions and interference. For example, in the antisaccade task (Hallett 1978; Hutchison 2007), subjects must inhibit the prepotent response of looking towards a flickering asterisk, and instead look in the opposite direction to detect a briefly presented letter. In the selective visual arrays task (Luck and Vogel 1997; Martin et al. 2021; Shipstead et al. 2014), subjects are shown a memory array and told to selectively attend to and remember a subset of items (e.g., remember the blue items) while ignoring the remaining items (e.g., ignore the red items). In the sustained attention to cue task (SACT, Draheim et al. 2021; Burgoyne et al. 2023; Draheim, Tshukara, and Engle 2023; Tsukahara and Engle 2023), subjects must remain focused on a cued spatial location on the computer screen for a variable wait period (2–12 s) to detect a briefly presented letter. These attention control measures have been tested extensively for individual differences research (see Burgoyne et al. 2023; Draheim et al. 2021; 2023; Kane et al. 2001; Martin et al. 2021; Redick, Heitz, and Engle 2007; Tsukahara and Engle 2023; Unsworth, Schrock, and Engle 2004; for details on their psychometric properties and evidence for their construct validity).

Antisaccade (Hallett 1978; Hutchison 2007). Participants identified a "Q" or "O" that appeared briefly on the opposite side of the screen as a distractor stimulus. After a central fixation cross appeared for 1000 ms or 2000 ms, an asterisk (*) flashed at an approximate 12.3° visual angle to the left or right of the central fixation for 100 ms. Afterward, the letter "Q" or "O" was presented on the opposite side at an approximate 12.3° visual angle from the central fixation for 100 ms, immediately followed by a visual mask (##). Participants indicated whether the letter was a "Q" or an "O". They completed 24 practice trials during which letter duration was set to 750 ms, followed by 72 test trials. The measure of performance was the proportion correct (i.e., minimum = 0%, maximum = 100%) and the task took approximately 12 min to administer.

Selective Visual Arrays (Luck and Vogel 1997; Martin et al. 2021; Shipstead et al. 2014). After a central fixation of 1000 ms, a cue word ("RED" or "BLUE") appeared instructing the participant to attend to either red or blue rectangles. Next, a target array of red and blue rectangles of different orientations (horizontal, left diagonal, right diagonal, and vertical) was presented for 250 ms, followed by a blank screen for 900 ms. Next, a probe array with only the cued-color rectangles was presented, with one rectangle highlighted by a white dot. The orientation of the highlighted rectangle was either the same as it was in the target array, or different, with equal likelihood. The participant indicated with the keyboard whether the orientation of the highlighted rectangle had changed or stayed the same. The target array contained either 5 or 7 rectangles per color (10 and 14 total). There were 40 trials per array set size. The measure was a capacity score (k), calculated using the single-probe correction (Cowan et al. 2005): set size * (hit rate + correction rejection rate − 1). The measure was the mean k estimate for the two set sizes (i.e., minimum = 0, maximum = 6) and the task took approximately 15 min to administer.

Sustained Attention to Cue (SACT; adapted from Draheim et al. 2021; Burgoyne et al. 2023; Draheim, Tshukara, and Engle 2023; Tsukahara and Engle 2023). Participants needed to sustain their attention on a visual circle cue presented at random locations on the screen and ultimately identify a target letter presented briefly at the center of the cue. Each trial started with a central black fixation. On half of the trials, the fixation was presented for 2 s and for the other half the fixation was presented for 3 s. After the fixation, following a 300 ms tone, a large white circle cue was presented in a randomly determined location on either the left or right side of the screen. To orient the participant to the circle cue, the large circle began to immediately shrink in size until it reached a fixed size. Once the cue reached the fixed size, after a variable wait time (equally distributed among 2, 4, 8, and 12 s), a white distracting asterisk appeared at the center of the screen. The asterisk blinked on and off in 100 ms intervals for a total duration of 300 ms (on for 100 ms, off for 100 ms, on for 100 ms). Then, a 3 ×3 array of letters was displayed at the center of the cue. The letters in the array consisted of B, D, P, and R. The central letter was the target letter and was presented in a dark gray font. The non-target letters were presented in black font with each letter occurring twice in the array and the target letter occurring three times. After 125 ms the central letter was masked with a # for 1,000 ms. Only the central target letter was masked. After the mask, the response options were displayed in boxes horizontally across the upper half of the screen. The participant used the mouse to select whether the target was a B, D, P, or R. Feedback was given during the practice trials but not the experimental trials. Sixty-four trials were administered. Accuracy rate was the dependent variable (i.e., minimum = 0%, maximum = 100%) and the task took approximately 15 min to administer.

2.6 | Fluid Intelligence

Raven's Advanced Progressive Matrices (Raven and Court 1998). Participants were presented with 3 × 3 arrays of geometric patterns. Each array contained a missing item, and participants were to select the pattern that best completed the array. Participants were given 10 min to complete the 18 odd-numbered items from Raven's Advanced Progressive Matrices. The measure was the number correct (i.e., minimum = 0, maximum = 18).

2.7 | Working Memory Capacity

Mental Counters (adapted from Alderton, Wolfe, and Larson 1997). This test challenged participants to keep track of three different values as they changed. Participants were presented with three lines in the center of the screen. On each trial, each line would begin with a value of five. Boxes would appear one at a time above or below the lines for 500 to 830 ms and then disappear, and the participant's task was to add "1" to that line's value if a box appeared above the line and subtract "1" from that line's value if a box appeared below the line. After a series of boxes, the participant was asked to report the value for each of the three lines. There were five trials at set size five (e.g., five boxes appeared during the trial), 14 trials at set size six, and 13 trials at set size seven, for a total of 32 trials. The measure of performance was the partial score, reflecting the number of counter values reported in the correct serial position. In other words, if a participant correctly reported two of the three counter values on a given trial, they would receive two points for that trial, out of a maximum of three points (i.e., minimum = 0, maximum = 96). The task took approximately 15 min to administer.

Advanced Rotation Span (Kane et al. 2004). Participants remembered a series of directional arrows (8 directions) of varying size (small or large) in alternation with a mental rotation task in which they had to mentally rotate and decide if a letter was mirror reversed or not. Set sizes ranged from 2 to 7 memory items and each set occurred 2 times. We used the partial score as the measure of performance, which awards points for however many memory items were recalled in the correct serial position (i.e., minimum = 0, maximum = 54). In other words, participants could earn partial credit on a trial if they recalled some of the items in the correct order. The task took approximately 15 min to administer.

2.8 | Military Selection Test Composite Scores

2.8.1 | Armed Forced Qualification Test (AFQT)

All enlisted military personnel (i.e., excluding military officers) completed the Armed Services Vocational Aptitude Battery (ASVAB) as part of the standard military enlistment process. Armed Forced Qualification Test (AFQT) composite scores were computed based on a linear average of the Mathematics Knowledge, Arithmetic Reasoning, Paragraph Comprehension, and Word Knowledge subtests.

2.8.2 | Aviation Selection Test Battery (ASTB)

All military officers (i.e., excluding enlisted personnel) participating in this study completed the Aviation Selection Test Battery (ASTB) as part of the selection process to be considered for pilot and flight officer (i.e., navigator) training programs. The ASTB comprises six subtests (i.e., Math Skills Test, Reading Comprehension Test, Mechanical Comprehension Test, Aviation and Nautical Information Test, Naval Aviation Trait Facet Inventory, and the Performance Based Measures Battery) (Navy Medicine n.d.).

2.8.3 | Academic Qualifications Rating Score (AQR Score)

All individuals who took the ASTB received an Academic Qualifications Rating (AQR) composite score, which is designed to predict academic performance during the following phases of flight school: Aviation Preflight Indoctrination (API) and Primary ground/academic school. This composite score was extracted based on performance on the ASTB subtests using the current (proprietary) weighting scheme used by the U.S. Navy. We used the stanine score as the measure of performance.

2.8.4 | Pilot Flight Aptitude Rating Score (PFAR Score)

All individuals who took the ASTB received a Pilot Flight Aptitude Rating (PFAR) composite score, which is designed to predict flight performance during the Primary phase of flight school for student naval aviators. A composite score was extracted based on performance on the ASTB subtests using the current (proprietary) weighting scheme used by the U.S. Navy. We used the stanine score as the measure of performance.

2.8.5 | Flight Officer Aptitude Rating Score (FOFAR Score)

All individuals who took the ASTB received a Flight Officer Aptitude Rating (FOFAR) composite score, which is designed to predict flight performance during the Primary phase of flight school for student naval flight officers. A composite score was extracted based on performance on the ASTB subtests using the current (proprietary) weighting scheme used by the U.S. Navy. We used the stanine score as the measure of performance.

Applicants for the Navy and Coast Guard pilot positions must achieve an AQR of four or higher and a PFAR of five or higher to meet minimum qualifications. Individuals applying to the Marine Corps pilot positions must achieve an AQR of five or higher and PFAR of six or higher to be selected. Applicants for the Navy and Coast Guard flight officer positions must achieve an AQR of four or higher and a FOFAR of five or higher to meet minimum qualifications. Individuals applying to the Marine Corps naval flight officer positions must achieve an AQR of five or higher and FOFAR of six or higher to be selected.

2.9 | Criterion Measures

2.9.1 | Number of Academic Setbacks

For the air traffic controllers in training, we counted the number of academic setbacks each student encountered. For context, the training program consists of multiple units of content, with some units subdivided into smaller courses lasting around 1 week. Students are tested regularly, and if they do not meet the academic criterion to move on, they are setback and must repeat that course or unit. The training typically takes around 14 weeks without any setbacks; academic setbacks are costly and to be avoided if possible.

2.9.2 | Academic Attrition

For the air traffic controllers in training, we collected data on whether they graduated the course or failed to graduate due to poor academic performance. For context, if an air traffic controller in training cannot meet the academic standard required by the training program (i.e., graduating with fewer than four academic setbacks), they are dropped from the program. This measure is a binary variable reflecting whether or not the trainee failed due to academic challenges.

2.9.3 | First Pass Pipeline Success

The measure "First Pass Pipeline Success" is a binary variable reflecting whether or not the air traffic controller in training was able to complete the training program without any academic setbacks. This is the Navy's primary dependent variable when evaluating which subtests should be used for each rating (i.e., job).

2.9.4 | Aviation Preflight Indoctrination Navy Standard Score (API NSS)

For student naval aviators and student naval flight officers, the Aviation Preflight Indoctrination Navy Standard Score (API NSS) represents a normalized summary of performance during the Aviation Preflight Indoctrination phase of flight school, which takes around 6 weeks to complete. During this phase, student naval aviators were introduced to flight basics within a classroom setting. Upon successful completion of this phase, student naval aviators and student naval flight officers transitioned to the primary phase of flight school, described next.

2.10 | Primary Academic Navy Standard Score (Primary Academic NSS)

For the student naval aviators and student naval flight officers, Primary Academic NSS represents a normalized summary of all classroom grades during the Primary phase of flight school. Primary flight training consists of learning the basics of aviation in the T-6 platform. The academic portions of training cover topics such as information on flight rules and regulations, flight navigation, aircraft systems knowledge, and flight operations and planning. The student naval aviators and student naval flight officers train separately, and their training has different objectives. The student naval flight officers get more training on navigation (i.e., we can think of them as "navigators"), while the student naval aviators have more emphasis on airmanship and flight maneuvering (i.e., we can think of them as "pilots").

2.10.1 | Primary Flight Navy Standard Score (Primary Flight NSS)

For the student naval aviators and student naval flight officers, Primary Flight NSS represents a normalized summary of all flight grades during the Primary phase of flight school. For the student naval aviators who will be responsible for flying the aircraft, they receive substantial training in the T-6B Texan II (see Figure 2) including multiple solo flights. For the student naval flight officers their primary training is done with the older T-6A Texan and is meant to familiarize them with the aircraft. The student naval flight officers' primary program includes some training flying the aircraft. However, this training is much less involved than the student naval aviators' training since the student naval flight officers will not be responsible for flying the aircraft once in the fleet. During this phase, student naval aviators were assessed on their performance flying a T-6B Texan II as well as in flight simulators.

2.11 | Transparency and Openness

We report all data exclusions below. This study's design and its analysis were not pre-registered. Data for this study are kept and protected by the U.S. Navy (specifically, the Naval Aerospace Medical Institute and the Naval Research Laboratory) and can only be shared if the requester can demonstrate to all relevant parties that required data security protocols will be

FIGURE 2 | T-6B Texan II Turboprop Trainer. U.S. Navy photo by Lt. Michelle Tucker; copyright: public domain.

adhered to. Analyses were conducted using IBM SPSS Statistics (IBM Corp 2021) primarily using the graphic-user-interface, however, the analysis code for the commonality analyses has been posted to the Open Science Framework (https://osf.io/a9842/?view_only=4bfb41c3c1db469480ba087e504c3dda).

2.12 | Data Preparation

We removed trainees' scores on a cognitive task if they showed severely poor performance indicating they did not understand the instructions or were not performing the task as intended. Specifically, we computed chance-level performance on each task; scores that were at or below chance-level performance were identified as problematic data points and set to missing. This procedure was applied to the three attention control tests (i.e., antisaccade, selective visual arrays, and the sustained attention to cue task; SACT). For the advanced rotation span task, problematic data points were defined by chance-level performance or worse on the processing subtask. We did not remove datapoints representing sub-chance performance on Raven's matrices or mental counters, as scores of zero are possible even if subjects understand the instructions. We then performed a two-pass outlier exclusion procedure on all tasks. We removed data points that were more than 3.5 standard deviations worse than the training-program sample mean two times, recomputing the sample mean and standard deviation each time.

2.13 | Restriction of Range

It is important to note that because our analyses of concurrent validity are estimated using incumbent samples, they are likely affected by restriction of range (Carretta and Ree 2022; Sackett and Yang 2000; Sackett et al. 2022; Schmidt, Hunter, and Urry 1976). Sailors were selected for these training programs based on a combination of scores, including AFQT scores to enlist in the military, multiple composites of ASVAB subtests (for the air traffic controllers in training), and for the aviators, Academic Qualification Rating performance, Pilot Flight Aptitude Rating Scores (for the student naval aviators), and Flight Officer Aptitude Rating Scores (for the student naval flight officers). Because sailors were selected directly based on selection test performance, the validities for these measures are likely attenuated due to direct range restriction (Carretta and Ree 2022). Additionally, the remaining measures are subject to indirect range restriction, to the extent that they are correlated with the selection test scores, which also attenuates validity estimates—although to a lesser degree than is the case for direct range restriction (Sackett et al. 2022). In other words, the observed validities we report are likely differentially affected due to direct range restriction on the selection measures and indirect range restriction on the other measures. We note that this puts the military selection tests at a disadvantage with respect to criterion-related validity when compared to the experimental cognitive ability measures, because of the differential effects of direct versus indirect range restriction.

Unfortunately, there were several challenges that prevented us from fully correcting for restriction of range in the present study because they would require us to make inferences and assumptions that go well beyond our data. Most importantly, we were unable to obtain data on unrestricted applicant samples, and therefore, we could not perform the multivariate correction for direct and indirect range restriction because it requires the unrestricted variances and covariances of the selection measures (Carretta and Ree 2022).

For the air traffic controllers in training, we were able to correct correlations for direct range restriction on AFQT scores and indirect range restriction on the other cognitive ability measures. We based our corrections on the unrestricted standard deviation of AFQT scores from the National Longitudinal Study of Youth 1997 data set (AFQT $SD = 29.17$; Bureau of Labor Statistics, U.S. Department of Labor 2019). We corrected for direct range restriction using formula #3 from Brown, Oswald, and Converse (2017) and corrected for indirect range restriction using formula #4; both formulas equivalent to those provided by Thorndike (1949). In the Results section, we perform a regression analyses predicting academic setbacks using both the observed and the corrected correlation matrices. We used the R function "lmCor" from the "psych" package (William Revelle 2024) to conduct regression analyses on the corrected correlation matrix. We did not perform logistic regression analyses on the corrected correlation matrix (i.e., predicting first pass pipeline success or academic attrition) because we were unaware of an established procedure for doing so.

For the student naval aviators and naval flight officers, selection consists of a multiple-hurdle design for which we were unable to obtain the requisite data to perform statistical corrections. Rather than potentially overcorrect the validity estimates in the presence of uncertainty, we elected to report the observed concurrent validity estimates in the incumbent samples while noting that these are likely underestimates of the validities that would be observed in unrestricted samples. Our decision is supported by Sackett et al. (2022), who state: "We reiterate our principle of conservative estimation: if one is not confident in the basis for a range restriction correction, it is better to forego a correction than to use a value that results in an overestimate. We suggest presenting the value obtained without the correction, and noting that as some degree of restriction is likely, the presented value is a conservative estimate" (pp. 60–61).

3 | Results

We present the results organized by occupational training group, beginning with the air traffic controllers, followed by the student naval aviators, and concluding with the student naval flight officers.

3.1 | Air Traffic Controllers

Demographic information for the 119 air traffic controllers in training is presented in Table 1. As shown in Table 2, the first pass pipeline success rate was fairly low (34%); a majority of trainees had one or more academic setbacks. Indeed, the average number of setbacks was 1.04 (SD = 1.03). The academic attrition rate was substantial, with 24.6% of trainees failing out of the air traffic control training program due to insufficient academic performance.

TABLE 1 | Demographic information for air traffic controllers in training.

Demographic	Statistic
Age (years)	Mean: 21.21 SD: 3.88 Range: 18–38 $N = 92$
At least some college?	Yes: 43% No: 57% $N = 94$
Race/ethnicity	White: 58% Black or African American: 25% Asian: 3% American Indian: 3% Hawaiian: 1% Hispanic: 24% $N = 118$

Note: Race/ethnicity values do not add to 100% because categories are not mutually exclusive. No gender information was reported for this portion of the sample.

Correlations between air traffic controller training performance, AFQT scores, and the cognitive ability measures are presented in Table 3. The attention control test selective visual arrays was a significant predictor of all air traffic controller training outcomes, demonstrating correlations that were moderate in magnitude in a down-selected sample of trainees. In terms of each criterion measure, first pass pipeline success correlated significantly with AFQT scores ($r = 0.26$, 95% CI [0.08, 0.42], $p = 0.005$), selective visual arrays ($r = 0.32$ [0.09, 0.52], $p = 0.008$), and mental counters ($r = 0.23$ [0.04, 0.41], $p = 0.021$). Number of academic setbacks correlated significantly with AFQT scores ($r = -0.26$ [−0.42, −0.08], $p = 0.006$) and selective visual arrays ($r = -0.30$ [−0.50, −0.06], $p = 0.016$). Academic attrition was significantly correlated with AFQT scores ($r = -0.22$ [−0.39, −0.04], $p = 0.018$) and selective visual arrays ($r = -0.40$ [−0.58, −0.17], $p = 0.001$).

Regarding the experimental predictor measures, the three tests of attention control (i.e., antisaccade, SACT, and selective visual arrays) correlated significantly with each other (rs ranged from 0.29 to 0.46) and with mental counters (rs ranged from 0.28 to 0.52). Although observed correlations between the AFQT and the experimental predictor measures were positive (rs ranged from 0.12 to 0.24), they were not statistically significant (all $p\text{s} \geq 0.06$), perhaps due to the smaller sample sizes for these correlations (median $n = 68$).

TABLE 2 | Descriptive statistics for air traffic controllers in training.

Measure	*N*	*M* or %	SD	Skew	Kurtosis	Reliability
First pass pipeline success	119	36.1%	—	0.59	−1.69	—
Academic setbacks	118	1.04	1.03	0.81	−0.06	—
Academic attrition	118	24.4%	—	1.20	−0.58	—
AFQT Score	116	72.90	11.60	−0.13	−0.09	—
Antisaccade	68	80.7%	11.4%	−0.55	−0.75	0.84[α]
SACT	71	86.1%	9.2%	−0.41	−0.83	0.87[α]
Selective visual arrays	66	2.14	1.11	0.30	−0.31	0.83[b]
Mental counters	98	74.78	12.64	−0.62	−0.08	0.89[α]

Note: First pass pipeline success and academic attrition are binary variables; we report the percentage of the sample who achieved first pass success or dropped out.
[α]Cronbach's alpha.
[b]Split-half reliability with Spearman-Brown correction.

TABLE 3 | Correlations for air traffic controllers in training.

Measure	1.	2.	3.	4.	5.	6.	7.	8.
1. First pass pipeline success	—	**−0.83**	**−0.56**	**0.56**	**0.33**	**0.24**	**0.49**	**0.39**
2. Academic setbacks	**−0.77**	—	**0.59**	**−0.56**	−0.18	**−0.20**	**−0.47**	**−0.33**
3. Academic attrition	**−0.43**	**0.47**	—	**−0.49**	**−0.36**	−0.20	**−0.53**	**−0.28**
4. AFQT score	**0.26**	**-0.26**	**−0.22**	—	0.40	0.29	0.53	0.44
5. Antisaccade	0.18	0.01	−0.23	0.17	—	**0.36**	**0.55**	**0.43**
6. SACT	0.12	−0.08	−0.09	0.12	**0.29**	—	**0.48**	**0.35**
7. Selective visual arrays	**0.32**	**−0.30**	**−0.40**	0.24	**0.46**	**0.41**	—	**0.61**
8. Mental counters	**0.23**	−0.16	−0.12	0.19	**0.33**	**0.28**	**0.52**	—

Note: The lower triangle represents observed correlations; the upper triangle represents correlations corrected for restriction of range (i.e., direct on AFQT, incidental on the other measures). Boldface, statistically significant at $p < 0.05$. Statistical significance of corrected correlations was determined via corrected 95% CIs. Pairwise *N* ranges from 60 to 118.

After correcting the correlations for direct range restriction on AFQT scores and indirect range restriction on the other measures, correlations with AFQT scores increased substantially, whereas correlations with the other experimental predictors increased to a lesser degree. The general trends were similar to those characterizing the observed correlation matrix. For instance, AFQT scores correlated $r_c = 0.56$, $r_c = -0.56$, and $r_c = -0.49$ with the three criterion measures, whereas selective visual arrays correlated $r_c = 0.49$, $r_c = -0.47$, and $r_c = -0.53$ with those criterion measures. Following the range restriction correction, many of the experimental measures became significant predictors of first pass pipeline success, academic setbacks, and academic attrition (see the upper triangle of the correlation matrix in Table 3).

We conducted a series of regression analyses predicting air traffic controller training performance. For each criterion measure, we examined (1) how much variance the attention control measures accounted for; (2) how much variance the AFQT accounted for; and (3) whether the attention control measures improved the prediction of training performance above and beyond the AFQT, using hierarchical regression analyses with AFQT entered in Step 1 and the attention control measures entered in Step 2. We also conducted commonality analyses to estimate the amount of unique and shared variance accounted for by the attention control measures and AFQT.

Results for the prediction of academic setbacks are presented in Table 4. Unfortunately, due to low statistical power, the overall model for the attention control measures was not statistically significant ($F(3, 54) = 2.10$, $p = .11$, $R = 0.323$, $R^2 = 10.4\%$), nor was the hierarchical regression model with AFQT entered in Step 1 and the attention control measures entered in Step 2 ($F(4, 51) = 2.36$, $p = 0.065$; $\Delta R = 0.106$, $\Delta R^2 = 7.3\%$, $\Delta p = .24$).

Next, we conducted a regression analysis predicting academic setbacks based on the corrected correlation matrix. As shown in Table 5, the model with just the attention control measures significantly predicted academic setbacks ($R = 0.48$, $R^2 = 0.23$, model $p = 0.002$); selective visual arrays was a significant predictor ($\beta = -0.54$, $p < .001$). For comparison, the model with just AFQT scores also significantly predicted academic setbacks ($R = 0.56$, $R^2 = 0.31$, model $p < 0.001$). Finally, we conducted a hierarchical regression analysis based on the corrected correlation matrix with AFQT scores entered in Step 1 and the attention control measures entered in Step 2. Although the inclusion of the attention control measures increased the model R by 0.08 and R^2 by 0.07, this change was not statistically significant ($\Delta p = 0.143$). Nevertheless, in the full model, both AFQT scores ($\beta = -0.46$, $p < .001$) and selective visual arrays ($\beta = -0.35$, $p = 0.027$) were significant predictors.

Results for the prediction of academic attrition are presented in Table 6. Academic attrition is a binary variable, so we used logistic regression and report classification accuracy and Nagelkerke's R^2. On their own, the attention control measures led to a classification accuracy of 81.0% (Nagelkerke $R^2 = 26.9\%$), and selective visual arrays was a significant predictor ($B = -1.30$,

TABLE 4 | Regression analyses predicting academic setbacks for air traffic controllers in training.

Model	Step	Model *df*	*F*	Model *p*	*R*	R^2	ΔR	ΔR^2	Δp	Measure	β	*p*
1	**1**	3, 54	2.10	0.111	0.323	0.104	—	—	—	Antisaccade	0.07	0.66
										SACT	0.14	0.33
										Visual Arrays	−0.37	0.018
2	**1**	1, 113	7.88	0.006	0.255	0.065	—	—	—	AFQT Scores	−0.26	0.006
3	**1**	1, 54	4.94	0.031	0.289	0.084	—	—	—	AFQT Scores	−0.25	0.071
	2	4, 51	2.36	0.065	0.395	0.156	0.106	0.073	0.236	Antisaccade	0.09	0.54
										SACT	0.16	0.26
										Visual Arrays	−0.29	0.056

Note: Parameter estimates (β and p) for the measures in Model 3 are reported for the full model, with all variables included in the regression equation.

TABLE 5 | Regression analyses predicting academic setbacks for air traffic controllers in training based on corrected correlation matrix.

Model	Step	Model *df*	*F*	Model *p*	*R*	R^2	ΔR	ΔR^2	Δp	Measure	β	*p*
1	**1**	3, 54	5.52	0.002	0.48	0.23	—	—	—	Antisaccade	0.12	0.43
										SACT	0.01	0.92
										Visual Arrays	−0.54	< 0.001
2	**1**	1, 113	51.86	< 0.001	0.56	0.31	—	—	—	AFQT Scores	−0.56	< 0.001
3	**1**	1, 54	24.78	< 0.001	0.56	0.31	—	—	—	AFQT Scores	−0.46	< 0.001
	2	4, 51	7.92	< 0.001	0.62	0.38	0.08	0.07	0.143	Antisaccade	0.18	0.18
										SACT	0.03	0.82
										Visual Arrays	−0.35	0.027

Note: Parameter estimates (β and p) for the measures in Model 3 are reported for the full model, with all variables included in the regression equation.

TABLE 6 | Logistic regression analyses predicting academic attrition for air traffic controllers in training.

Model	Step	df	χ^2	Model p	Model n	Nagel. R^2	Class. Acc.	ΔClass. Acc.	Δp	Measure	B	SE	p
1	1	3	11.24	0.011	58	0.269	81.0%	—	—	Antisaccade	1.39	3.51	0.69
										SACT	4.42	4.43	0.32
										Visual Arrays	−1.30	0.47	0.006
2	1	1	5.71	0.017	115	0.072	74.8%	—	—	AFQT Scores	−0.05	0.02	0.021
3	1	1	6.62	0.010	56	0.169	76.8%	—	—	AFQT Scores	−0.08	0.04	0.037
	2	4	17.38	0.002	56	0.403	82.1%	5.30%	0.013	Antisaccade	2.36	3.59	0.51
										SACT	4.49	5.07	0.38
										Visual Arrays	−1.58	0.60	0.008

Note: Parameter estimates for the measures in Model 3 are reported for the full model, with all variables included in the regression equation.
Abbreviations: Class. Acc. = classification accuracy, Nagel. = Nagelkerke R^2.

$SE = 0.47$, $p = 0.006$). For comparison, on its own, the AFQT led to a classification accuracy of 74.8% (Nagelkerke $R^2 = 7.2\%$).

Next, we tested whether the attention control measures predicted academic attrition above and beyond AFQT scores. In Step 1, AFQT scores led to a classification accuracy of 76.8% (Nagelkerke $R^2 = 16.9\%$). In Step 2, including the attention control measures led to a classification accuracy of 82.1% (Nagelkerke $R^2 = 40.3\%$). The improvement in prediction—an increase of 5.3% in classification accuracy and 23.4% in terms of Nagelkerke's R^2—was statistically significant (χ^2 $(3) = 10.75$, $p = 0.013$).

Results for the logistic regression analyses predicting first-pass pipeline success are presented in Table 7. The model with just the attention control measures was not statistically significant ($p = 0.122$). By contrast, on its own, the AFQT led to a classification accuracy of 70.7% (Nagelkerke $R^2 = 9.3\%$) and the model was statistically significant ($p = 0.004$).

We tested whether the attention control measures predicted first pass pipeline success above and beyond AFQT scores. In Step 1, AFQT scores led to a classification accuracy of 69.6% (Nagelkerke $R^2 = 17.6\%$). In Step 2, including the attention control measures increased the classification accuracy to 71.4% (Nagelkerke $R^2 = 26.2\%$), however, the improvement in prediction was not statistically significant ($p = 0.232$).

Finally, we performed a commonality analysis, which builds on the principles of linear regression and allowed us to estimate the proportion of variance in academic setbacks that was uniquely accounted for by AFQT scores and attention control, as well as the common variance that was accounted for by both AFQT scores and attention control. We did not perform commonality analysis on the dichotomous criterion measures because this technique is based on linear regression, not logistic regression. We used the SPSS syntax provided by Nimon (2010). As depicted in Figure 3, the commonality analysis revealed that when predicting academic setbacks, AFQT scores accounted for 5.19% of the unique variance, attention control accounted for 9.10% of the unique variance, and AFQT and attention control shared 1.33% of the variance that they accounted for.

3.2 | Student Naval Aviators

Demographic information for the 293 student naval aviators is presented in Table 8; descriptive statistics are presented in Table 9.

Correlations between student naval aviators' training performance, the composite scores AQR (i.e., Academic Qualifications Rating) and PFAR (i.e., Pilot Flight Aptitude Rating), and the cognitive ability measures are presented in Table 10. To reiterate, there are three criterion measures of pilots' training performance considered here: API NSS represents performance in the aviation preflight indoctrination portion of training, Primary Flight NSS represents performance in the flight portion of primary training, and Primary Academic NSS represents performance in the academic portion of primary training. AQR stanine was a significant predictor of all three criterion

TABLE 7 | Logistic regression analyses predicting first pass pipeline success for air traffic controllers in training.

Model	Step	*df*	χ^2	Model *p*	Model *n*	Nagel. R^2	Class. Acc.	ΔClass. Acc.	Δ*p*	Measure	*B*	*SE*	*p*
1	**1**	3	5.80	0.122	58	0.128	63.8%	—	—	Antisaccade	1.16	2.97	0.70
										SACT	−3.11	3.71	0.40
										Visual Arrays	0.62	0.31	0.045
2	**1**	1	8.18	0.004	116	0.093	70.7%	—	—	AFQT Scores	0.05	0.02	0.007
3	**1**	1	7.87	0.005	56	0.176	69.6%	—	—	AFQT Scores	0.07	0.03	0.023
	2	4	12.15	0.016	56	0.262	71.4%	1.80%	.232	Antisaccade	0.86	3.22	0.79
										SACT	−3.99	3.97	0.32
										Visual Arrays	0.62	0.35	0.075

Note: Parameter estimates for the measures in Model 3 are reported for the full model, with all variables included in the regression equation.
Abbreviations: Class. Acc. = classification accuracy; Nagel. = Nagelkerke R^2.

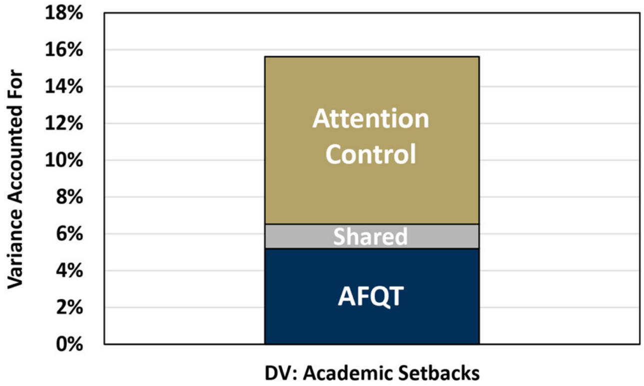

FIGURE 3 | Results of the commonality analysis predicting academic setbacks for air traffic controllers in training. Note that these results are based on observed correlations, not correlations corrected for restriction of range.

TABLE 8 | Demographic information for student naval aviators in training.

Demographic	Statistic
Age (years)	Mean: 23.92 SD: 1.96 Range: 21–31 *N* = 246
Sex	Male: 89.1% Female: 10.9% *N* = 293
Race/ethnicity	White: 91.8% Black or African American: 3.8% Asian: 5.8% American Indian: 1.4% Hawaiian: 1.0% Hispanic: 0.7% *N* = 293

Note: Race/ethnicity values do not add to 100% because categories are not mutually exclusive. No education information was reported for this portion of the sample.

measures, API NSS ($r = 0.36$, 95% CI [0.25, 0.46], $p < 0.001$), Primary Flight NSS ($r = 0.16$ [0.04, 0.27], $p = 0.012$), and Primary Academic NSS ($r = 0.20$ [0.08, 0.31], $p = .001$). For comparison, PFAR stanine was a significant predictor of Primary Flight NSS ($r = 0.17$ [0.05, 0.29], $p = 0.006$).

Unfortunately, none of the experimental predictor measures were significantly correlated with Student Naval Aviator training performance; the strongest relation was between the attention control test antisaccade and Primary Flight NSS ($r = 0.13$ [0.00, 0.26], $p = 0.050$). Incremental validity analyses were nonsignificant for this sample and are reported in the Supplemental Materials.

3.3 | Student Naval Flight Officers

Demographic information for the 78 student naval flight officers is presented in Table 11; descriptive statistics are presented in Table 12.

TABLE 9 | Descriptive statistics for student naval aviators in training.

Measure	N	M	SD	Skew	Kurtosis	Reliability
API NSS	267	49.42	6.28	0.01	−0.56	—
Primary flight NSS	262	49.44	9.73	0.03	0.12	—
Primary academic NSS	264	47.06	10.07	−0.28	−0.44	—
AQR stanine	289	6.58	1.13	0.00	−0.39	—
PFAR stanine	289	6.91	0.88	−0.39	−0.40	—
Antisaccade	252	83.5%	9.6%	−1.11	0.87	0.80[α]
SACT	245	92.6%	6.1%	−1.43	2.43	0.64[α]
Selective visual arrays	228	2.32	1.04	0.09	-0.59	0.74[b]
Mental counters	154	79.16	10.01	−0.65	0.16	0.80[α]
Advanced rotation span	100	24.95	7.93	−0.38	0.03	0.73[α]
Raven's matrices	102	11.34	2.57	−0.45	0.19	0.54[b]

[α]Cronbach's alpha.
[b]Split-half reliability with Spearman-Brown correction.

TABLE 10 | Correlations for student naval aviators in training.

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. API NSS	—									
2. Primary fight NSS	**0.36**	—								
3. Primary academic NSS	**0.53**	**0.38**	—							
4. AQR stanine	**0.36**	**0.16**	**0.20**	—						
5. PFAR stanine	0.02	**0.17**	0.00	**0.39**	—					
6. Antisaccade	−0.03	0.13	0.03	0.05	0.10	—				
7. SACT	0.05	0.06	0.01	0.07	0.03	**0.25**	—			
8. S. Visual arrays	0.05	−0.01	−0.08	**0.22**	0.01	**0.29**	**0.16**	—		
9. Mental counters	0.03	0.04	−0.09	0.12	0.05	0.14	0.13	**0.39**	—	
10. Rotation span	0.07	0.05	−0.14	−0.05	0.00	**0.24**	**0.21**	**0.24**	N/A	—
11. Raven's matrices	0.12	−0.01	0.01	**0.34**	0.06	0.00	0.01	**0.31**	N/A	**0.31**

Note: Boldface, statistically significant at $p < 0.05$. N/A = no correlation to report; student naval aviators who completed mental counters did not complete rotation span or Raven's matrices. Pairwise N ranges from 88 to 289.

Correlations between student naval flight officer training performance, AQR (i.e., academic qualifications rating) and FOFAR (i.e., flight officer flight aptitude rating), and the cognitive ability measures are presented in Table 13. API NSS correlated significantly with AQR stanine ($r = 0.30$, 95% CI [0.08, 0.49], $p = 0.009$), FOFAR stanine ($r = 0.27$ [0.05, 0.47], $p = 0.018$), and the attention control tests SACT ($r = 0.34$ [.10, 0.55], $p = 0.007$) and selective visual arrays ($r = 0.26$ [0.01, 0.49], $p = 0.045$). Primary Flight NSS and Primary Academic NSS did not correlate significantly with any of the other measures, although we note that sample sizes were small (pairwise ns ranged from 4 to 74, average $n = 32$).

We conducted regression analyses predicting API NSS (Table 14) and Primary Flight NSS (Table 15). We did not run analyses predicting Primary Academic NSS as there was an insufficient sample size ($N \leq 14$).

The model with just the attention control measures predicting API NSS was not statistically significant ($F(3, 49) = 2.69$, $p = 0.057$; $R^2 = 14.1\%$, $R = 0.376$). For comparison, AQR stanine accounted for 8.9% of the variance in API NSS on its own ($R = 0.298$) and the model was significant ($F(1, 74) = 7.22$, $p = 0.009$).

We tested whether the attention control measures predicted API NSS above and beyond AQR stanine. In Step 1, AQR stanine accounted for 13.0% of the variance ($R = 0.360$). In Step 2, the inclusion of the attention control measures increased the total variance explained to 20.7% ($R = 0.455$)—the change in R and R^2, although large, was not statistically significant ($\Delta R = 0.095$, $\Delta R^2 = 7.7\%$, $p = 0.22$), likely due to the small sample size.

Results for the prediction of Primary Flight NSS are presented in Table 15. For context, during this phase of training, student naval flight officers are trained on situational awareness, visual scanning, and procedural recall and execution. When predicting Primary Flight NSS, the model with just the attention control measures was not statistically significant ($F(3, 48) = 0.74$,

$p = .53$; $R^2 = 4.4\%$, $R = 0.210$). For comparison, the model with just FOFAR stanine was also not statistically significant ($F(1, 72) = 0.70$, $p = 0.41$; $R^2 = 1.0\%$, $R = 0.098$).

We tested whether the attention control measures predicted Primary Flight NSS above and beyond FOFAR stanine. In Step 1, FOFAR stanine accounted for 1.1% of the variance ($R = 0.106$). In Step 2, the inclusion of the attention control measures increased the total variance accounted for to 5.2% ($R = 0.229$)—the change in R and R^2 was not statistically significant ($\Delta R = 0.123$, $\Delta R^2 = 4.1\%$, $p = 0.59$), nor was the overall model ($p = 0.64$).

As depicted in Figure 4, the commonality analysis of API NSS (i.e., Aviation Preflight Indoctrination Navy Standard Score) revealed that AQR stanine (i.e., Academic Qualifications Rating Stanine) accounted for 6.58% of the unique variance, attention control accounted for 11.80% of the unique variance, and AQR and attention control shared 2.32% of the variance that they accounted for. When predicting Primary Flight NSS, FOFAR stanine (i.e., Flight Officer Aptitude Rating Stanine) accounted for 0.82% of the unique variance, attention control accounted for 4.27% of the unique variance, and FOFAR and attention control shared 0.14% of the variance that they accounted for.

TABLE 11 | Demographic information for student naval flight officers.

Demographic	Statistic
Age (years)	Mean: 23.16 SD: 1.73 Range: 21–31 $N = 62$
Sex	Male: 74.4% Female: 25.6% $N = 78$
Race/ethnicity	White: 83.3% Black or African American: 6.4% Asian: 14.1% American Indian: 2.6% Hawaiian: 6.4% Hispanic: 2.6% $N = 78$

Note: Race/ethnicity values do not add to 100% because categories are not mutually exclusive.

3.4 | Subgroup Differences

In our final set of analyses, we examined subgroup differences in performance on the measures administered to the air traffic controllers, student naval aviators, and student naval flight officers. We grouped samples together to maximize statistical power, and then examined relative differences in performance on training metrics, composite scores including the AFQT, AQR, PFAR, FOFAR, and cognitive ability measures as a function of age, sex, and race and ethnicity. The purpose of these analyses was to investigate whether tests of more fluid cognitive abilities, and in particular, tests of attention control, demonstrate smaller subgroup differences than tests of more crystallized cognitive abilities, such as the AFQT. Predictors that minimize group differences can help reduce adverse impact.

First, we examined whether there were differences in performance as a function of age. As shown in Table 16, age was significantly negatively correlated with Primary Flight NSS for student naval aviators and student naval flight officers ($r = -0.13$, $p = 0.026$), suggesting that younger trainees tended to perform slightly better. Age was positively correlated with AFQT ($r = 0.33$, $p = .002$) and performance on the sustained attention to cue task (i.e., SACT, $r = 0.28$, $p < 0.001$), indicating a slight advantage for somewhat older trainees.

We also examined performance differences between men and women (Table 17). We take caution in interpreting these results, as the sample of women was small (average $n = 40$). Four measures showed significant sex differences, favoring males: Primary Flight NSS ($d = 0.41$, $p = 0.009$), AQR stanine

TABLE 12 | Descriptive statistics for student naval flight officers.

Measure	N	M	SD	Skew	Kurtosis	Reliability
API NSS	77	47.44	6.74	0.23	−0.69	—
Primary flight NSS	75	52.68	8.70	−0.57	0.96	—
Primary academic NSS	15	44.85	10.81	−1.03	1.08	—
AQR stanine	77	6.49	1.06	0.46	−0.37	—
FOFAR stanine	77	6.70	1.05	−0.34	−0.50	—
Antisaccade	63	79.4%	12.1%	−0.67	−0.60	0.86[α]
SACT	61	91.7%	8.1%	−1.49	1.61	0.81[α]
Selective visual arrays	59	2.15	1.08	−0.10	−0.76	0.77[b]
Mental counters	16	81.06	9.14	−1.71	4.61	0.82[α]
Advanced rotation span	48	24.31	7.14	0.03	−0.79	0.61[α]
Raven's matrices	46	10.98	2.57	−0.23	−0.16	0.78[b]

[α]Cronbach's alpha.
[b]Split-half reliability with Spearman-Brown correction.

TABLE 13 | Correlations for student naval flight officers.

Measure	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. API NSS	—									
2. Primary flight NSS	**0.49**	—								
3. Primary academic NSS	0.37	0.35	—							
4. AQR stanine	**0.30**	0.13	−0.15	—						
5. FOFAR stanine	**0.27**	0.10	−0.31	**0.79**	—					
6. Antisaccade	0.09	0.06	−0.07	−0.01	0.07	—				
7. SACT	**0.34**	0.16	0.33	0.18	0.18	0.20	—			
8. S. Visual arrays	**0.26**	0.05	0.34	0.22	0.16	**0.33**	**0.28**	—		
9. Mental counters	−0.17	0.14	−0.44	0.04	0.40	0.20	−0.20	0.02	—	
10. Rotation span	0.14	0.10	−0.04	**0.36**	0.24	−0.13	0.13	**r0.41**	N/A	—
11. Raven's matrices	0.00	0.21	0.14	0.19	0.09	−0.06	0.14	0.21	N/A	0.25

Note: Boldface, statistically significant at $p < 0.05$. N/A = no correlation to report; naval flight officers in training who completed mental counters did not complete rotation span or Raven's matrices. Pairwise N ranges from 42 to 76 for all measures except primary academic NSS and mental counters. For Primary Academic NSS, pairwise N ranges from 4 to 15; for mental counters, pairwise N ranges from 4 to 16.

TABLE 14 | Regression analyses predicting API NSS for student naval flight officers.

Model	Step	df	F	Model p	R	R^2	ΔR	ΔR^2	Δp	Measure	β	p
1	**1**	3, 49	2.69	0.057	0.376	0.141	—	—	—	Antisaccade	−0.01	0.97
										SACT	0.24	0.088
										Visual Arrays	0.23	0.11
2	**1**	1, 74	7.22	0.009	0.298	0.089	—	—	—	AQR Stanine	0.30	0.009
3	**1**	1, 50	7.46	0.009	0.360	0.130	—	—	—	AQR Stanine	0.27	0.055
	2	4, 47	3.07	0.025	0.455	0.207	0.095	0.077	0.220	Antisaccade	−0.01	0.95
										SACT	0.21	0.13
										Visual arrays	0.15	0.28

Note: Parameter estimates for the measures in Model 3 are reported for the full model, with all variables included in the regression equation.

TABLE 15 | Regression analyses predicting primary fight NSS for student naval flight officers.

Model	Step	df	F	Model p	R	R^2	ΔR	ΔR^2	Δp	Measure	β	p
1	**1**	3, 48	0.74	0.53	0.210	0.044	—	—	—	Antisaccade	0.11	0.46
										SACT	0.17	0.25
										Visual Arrays	−0.06	0.71
2	**1**	1, 72	0.70	0.41	0.098	0.010	—	—	—	FOFAR Stanine	0.10	0.41
3	**1**	1, 49	0.55	0.46	0.106	0.011	—	—	—	FOFAR Stanine	0.08	0.60
	2	4, 46	0.64	0.64	0.229	0.052	0.123	0.041	0.58	Antisaccade	0.06	0.71
										SACT	0.20	0.21
										Visual Arrays	−0.10	0.54

Note: Parameter estimates for the measures in Model 3 are reported for the full model, with all variables included in the regression equation.

($d = 0.51$, $p < 0.001$), PFAR stanine ($d = 0.67$, $p < 0.001$), and antisaccade ($d = 0.42$, $p = 0.010$).

Finally, we investigated performance differences across racial and ethnic subgroups (Table 18). We initially conducted analyses on ethnicity to compare Hispanic and Non-Hispanic participants, but the Hispanic sample sizes were as low as $n = 1$ for three measures (Academic NSS, rotation span, Raven's matrices) and the average sample size was $n = 14$ (the median was $n = 13$). Due to these small samples, we elected not to perform comparisons across these groups, however, the raw values are included in the Supplemental Materials to facilitate future meta-analytic work.

For the analyses of racial subgroups, we compared individuals who indicated they were White or some

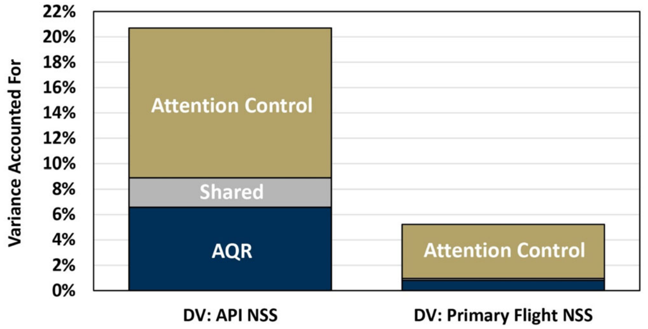

FIGURE 4 | Results of the commonality analysis predicting API NSS and Primary Flight NSS for naval flight officers. Note that for the dependent variable (DV) Primary Flight NSS, the blue-shaded region represents variance uniquely accounted for by FOFAR scores.

TABLE 16 | Correlations between age and performance and ability measures.

Measure	*r* with age	*p*	*n*
ATC first pass pipeline success	$r = -0.05$	$p = 0.64$	92
ATC setbacks	$r = -0.02$	$p = 0.84$	91
ATC attrition	$r = 0.06$	$p = 0.58$	91
Aviation preflight indoctrination NSS	$r = -0.04$	$p = 0.51$	285
Primary flight NSS	$\boldsymbol{r = -0.13}$	$\boldsymbol{p = 0.026}$	278
Primary academic NSS	$r = 0.02$	$p = 0.80$	230
AFQT score	$\boldsymbol{r = 0.33}$	$\boldsymbol{p = 0.002}$	90
AQR stanine	$r = -0.09$	$p = 0.13$	303
PFAR stanine	$r = 0.09$	$p = 0.11$	303
FOFAR stanine	$r = 0.02$	$p = 0.75$	303
Antisaccade	$r = 0.08$	$p = 0.12$	354
Sustained attention to cue task (SACT)	$\boldsymbol{r = 0.28}$	$\boldsymbol{p < 0.001}$	363
Visual arrays	$r = 0.00$	$p = 0.99$	335
Mental counters	$r = 0.01$	$p = 0.93$	253
Rotation span	$r = -0.01$	$p = 0.90$	140
Raven's matrices	$r = -0.13$	$p = 0.12$	142

Note: Bold, $p < 0.05$.

combination of White and other racial identities to individuals who indicated they were not White (specifically, those who selected Black, Hawaiian, and/or American Indian). This grouping served to maximize statistical power. We found significant differences between the two groups on AFQT scores ($d = 0.48$, $p = 0.014$), AQR Stanine ($d = 0.59$, $p = 0.009$), SACT ($d = 0.75$, $p < 0.001$), and mental counters ($d = 0.67$, $p < 0.001$). For the other two attention control measures, antisaccade and selective visual arrays, differences were close to zero and not statistically significant ($d = 0.19$ and $d = 0.15$, $ps > 0.31$). This indicates that some attention control measures may be better suited for reducing adverse impact than others.

4 | Discussion

Improving the prediction of occupational performance and training success is a critical goal for the U.S. military. Air traffic control and aviation are challenging occupational pursuits because they require the ability to maintain focus amidst distractions and interference. Furthermore, they entail a great deal of cognitive and motor skill learning (e.g., operating control panels, flying planes, communicating mission-critical information in a fast-paced environment). All these requirements point to an important role of attention control, but except for some subtests in the Aviation Selection Test Battery (ASTB), current selection tests such as the Armed Forces Qualification Test (AFQT) primarily measure acquired knowledge. Because crystallized knowledge tests may yield large group differences and leave room for improvement in the prediction of performance, we investigated whether tests of attention control could add incremental validity.

Commonality analyses revealed that attention control more than doubled the prediction of training performance (i.e., academic setbacks) for air traffic controllers in training. Specifically, attention control measures accounted for 9.1% of the unique variance in academic setbacks, whereas the AFQT accounted for 5.2% of the unique variance. There was little overlap in the variance that attention control and AFQT scores accounted for (1.3%). As a point of comparison, recall that Held, Carretta, and Rumsey (2014) estimated that an incremental validity of just $r = 0.02$ would lead to 15 fewer of every 1000 air traffic controller trainees failing, saving the Navy \$1.5 million per 1000 trainees. It follows that there may be considerable savings associated with augmenting current military selection tests with tests of attention control.

For the prediction of air traffic control academic attrition, attention control measures significantly increased classification accuracy above and beyond the AFQT, from 76.8% ($R^2_{\text{Nagelkerke}} = 16.9\%$) to 82.1% ($\Delta R^2_{\text{Nagelkerke}} = 40.3\%$). This suggests that attention control measures capture something important to air traffic control training success that the AFQT does not. From a theoretical standpoint, attention control reflects the ability to maintain focus on task-relevant information and resist distraction (Burgoyne and Engle 2020)—these skills would appear important for air traffic controllers both in the classroom and behind the control panel, as they are challenged to acquire new skills during training and consistently execute them while on the job.

For the student naval flight officers, hierarchical regression analyses revealed that attention control accounted for 10.4% of the incremental variance in API NSS (i.e., aviation preflight indoctrination training performance) above and beyond the military composite AQR (i.e., academic qualifications rating), and 2.1% of the incremental variance in flight performance above and beyond the military composite FOFAR score, although these increments fell shy of statistical significance due to low statistical power (i.e., small samples). Commonality analyses revealed that attention control accounted for 11.80% of the unique variance in API performance and 4.27% of the variance in flight performance, whereas the military composites accounted for 6.58% and 0.82% of the unique variance, respectively.

TABLE 17 | Standardized mean differences between men and women.

Measure	Male			Female			Cohen's *d*	*p* value
	M	SD	*n*	*M*	SD	*n*		
Aviation preflight indoctrination NSS	49.18	6.37	295	47.71	6.69	49	0.23	0.14
Primary flight NSS	50.72	9.70	288	46.87	8.24	49	**0.41**	**0.009**
Primary academic NSS	46.92	10.30	244	47.04	8.74	35	−0.01	0.95
AQR stanine	6.64	1.12	315	6.08	1.02	51	**0.51**	**< 0.001**
PFAR stanine	6.95	0.89	315	6.35	0.96	51	**0.67**	**< 0.001**
FOFAR stanine	6.79	0.96	315	6.63	0.92	51	0.17	0.25
Antisaccade	0.83	0.10	271	0.79	0.11	44	**0.42**	**0.010**
Sustained attention to cue task	0.92	0.06	266	0.92	0.08	40	0.08	0.62
Selective visual arrays	2.30	1.07	248	2.17	0.88	39	0.12	0.48
Mental counters	79.60	9.85	151	77.26	10.50	19	0.24	0.33
Rotation span	24.98	7.90	125	23.43	6.21	23	0.20	0.37
Raven's matrices	11.28	2.63	125	10.96	2.20	23	0.13	0.58

Note: Sex differences were not computed for air traffic controller in training first pass pipeline success, academics setbacks, academic attrition, or AFQT scores because sex data were not available for these variables. Positive Cohen's *d* values indicate that the mean was greater for men than for women.

TABLE 18 | Standardized mean differences across race.

Measure	White			Nonwhite			Cohen's *d*/% difference	*p* value
	M (%)	SD	*n*	*M* (%)	SD	*n*		
ATC first pass pipeline success	38%	—	69	34%	—	47	4%	0.69
ATC setbacks	1.01	1.03	68	1.09	1.06	47	-0.07	0.72
ATC attrition	26%	—	68	21%	—	47	5%	0.53
Aviation preflight Indoctrination NSS	49.17	6.40	309	46.84	5.92	19	0.37	0.12
Primary flight NSS	50.45	9.33	304	47.82	10.86	18	0.28	0.25
Primary academic NSS	46.97	10.29	255	44.64	6.18	13	0.23	0.22
AFQT score	74.87	11.06	69	69.43	11.81	44	**0.48**	**0.015**
AQR stanine	6.60	1.10	330	5.95	1.12	21	**0.59**	**0.009**
PFAR stanine	6.89	0.90	330	6.71	1.10	21	0.19	0.40
FOFAR stanine	6.78	0.95	330	6.48	1.03	21	0.32	0.16
Antisaccade	0.83	0.10	325	0.81	0.12	44	0.19	0.31
Sustained attention to cue task	0.92	0.07	315	0.86	0.10	48	**0.75**	**< 0.001**
Selective visual arrays	2.27	1.03	297	2.11	1.13	43	0.15	0.36
Mental counters	78.95	10.38	214	71.67	12.92	46	**0.67**	**< 0.001**
Rotation span	24.75	7.97	131	25.10	4.98	10	−0.05	0.89
Raven's matrices	11.35	2.56	131	10.50	2.42	10	0.33	0.31

Note: Cohen's *d* and associated *p* values are based on comparisons to the White subsample. Positive Cohen's *d* values indicate that the mean was greater for the White subsample than for the non-White subsample.

We considered two other cognitive ability constructs in our analyses: fluid intelligence and working memory capacity. We found that measures of working memory capacity (i.e., mental counters and rotation span) and fluid intelligence (i.e., Raven's matrices) did not account for statistically significant variance in training outcomes above and beyond current selection tests and attention control, although in some cases incremental validities were promising and would be of practical significance if replicated in larger samples (see the Supplemental Materials). At the bivariate level, mental counters correlated significantly with air traffic controllers' first pass pipeline success ($r = 0.23$); all other correlations were not significant.

Our analyses of subgroup differences revealed several interesting patterns. First, age was negatively correlated with aviators' flight performance and positively associated with performance on the AFQT and the Sustained Attention to Cue Task (SACT), despite the fact that the age range was relatively narrow across samples (i.e., 18–38). Second, males performed significantly better than females on two military tests (AQR and PFAR), flight performance, and the antisaccade test of attention control. Finally, four measures produced significant differences across White versus nonwhite racial groups: AFQT scores ($d = 0.48$, $p = 0.014$), AQR Stanine ($d = 0.59$, $p = 0.009$), SACT ($d = 0.75$, $p < 0.001$), and mental counters ($d = 0.67$, $p < 0.001$). By comparison, the other two attention control measures (i.e., antisaccade and selective visual arrays) showed group differences that were close to zero and nonsignificant ($ds \leq 0.19$). As we noted, evidence suggests that tests of accultured knowledge tend to lead to larger group differences than tests of more fluid abilities. Tests of attention control are relatively culture-free and do not contain much material taught in schools, unlike the AFQT, which taps math and verbal abilities. Thus, evidence suggests that tests of attention control (except for the SACT) could be a viable pathway towards reducing adverse impact in personnel selection (for additional examples, see Burgoyne, Mashburn, and Engle 2021 and Bosco, Allen, and Singh 2015). It is an open question for future research what explains the large difference in performance across groups on the SACT.

4.1 | Limitations

We note three limitations of the present work. First, our samples were small, which increased the uncertainty associated with our point estimates and reduced our power to detect significant effects. As even slight improvements in predictive validity can confer significant advantages to organizations (Held, Carretta, and Rumsey 2014), having a much larger sample to detect smaller effects would benefit future research. Because our sample included real-world student naval aviators, air traffic controllers, and student naval flight officers in training, acquiring a larger sample is not a trivial undertaking. Above and beyond the challenges of obtaining data on real-world sailors, there are also temporal challenges: Sailors were recruited to complete these cognitive ability tests towards the beginning of their training. As a result, it typically has taken more than 1 year to obtain data on their performance over the course of the training program. That said, we see this work as a step in the right direction, and hope that it will lead to data collection involving larger samples in the future. Additionally, we have sought to provide enough information about our results so that meta-analysts can include our studies as additional relevant data becomes available.

The second limitation is that we had limited testing time to administer our task battery to the student air traffic controllers, student naval aviators, and student naval flight officers. As a result, we prioritized the measurement of attention control. With more testing time, we could have included more measures of fluid intelligence and working memory capacity, added more trials to the tasks, and allowed more testing time for Raven's Matrices. Adding more trials, testing time, or longer wait intervals to the SACT would also be worthwhile, as the student naval aviators and naval flight officers were approaching ceiling-level performance on this task, which could have resulted in range compression and validity attenuation.

The third limitation is that we used an incumbent sample of trainees which was affected by direct range restriction on the selection measures and indirect range restriction on the other cognitive ability measures. Restriction of range leads to the attenuation of validity estimates (Carretta and Ree 2022; Sackett and Yang 2000; Sackett et al. 2022; Schmidt, Hunter, and Urry 1976). Importantly, direct range restriction has a greater impact than indirect range restriction (Sackett et al. 2022), and so it is likely that the observed validity estimates presented here are underestimates relative to what might be observed in an unrestricted applicant sample, particularly for the selection measures. Although we were able to correct correlations for the sample of air traffic controllers for direct and indirect range restriction, we did not have the requisite data to perform such corrections on the other samples.

5 | Conclusion

We investigated whether attention control tests could improve the prediction of training performance above and beyond current military selection tests in a sample of 490 air traffic controllers, student naval aviators, and student naval flight officers in training. Our results provide preliminary evidence for the validity of attention control measures for occupational training success.

Acknowledgments

This work was supported by a grant from the Naval Research Laboratory N00173-20-2-C003 to Randall W. Engle and a grant from the Office of Naval Research N00014-23-1-2768 to Randall W. Engle and Alexander P. Burgoyne. This study's design and its analysis were not pre-registered. Data for this study are kept and protected by the U.S. Navy (specifically, the Naval Aerospace Medical Institute and the rNaval Research Laboratory) and can only be shared if the requester can demonstrate to all relevant parties that required data security protocols will be adhered to. The analysis code for the commonality analyses has been posted to the Open Science Framework (https://osf.io/a9842).

Data Availability Statement

Data for this study are kept and protected by the U.S. Navy (specifically, the Naval Aerospace Medical Institute and the Naval Research Laboratory) and can only be shared if the requester can demonstrate to all relevant parties that required data security protocols will be adhered to. The analysis code for the commonality analyses has been posted to the Open Science Framework (https://osf.io/a9842).

General Audience Summary

Individual differences in the ability to control attention are correlated with a wide range of important outcomes, from cognitive task performance and academic achievement to health behaviors and emotion regulation. Nevertheless, current selection tests and composite scores used by the U.S. Navy to identify personnel for high-skill jobs primarily reflect individual differences in acquired knowledge, leaving room for improvement with respect to predictive validity and subgroup differences. In a sample of 490 U.S. Navy trainees, we found preliminary evidence that measures of attention control improve the prediction of training performance above and beyond current selection test scores.

References

Ahmed, S. F., S. Tang, N. E. Waters, and P. Davis-Kean. 2019. "Executive Function and Academic Achievement: Longitudinal Relations From Early Childhood to Adolescence." *Journal of Educational Psychology* 111, no. 3: 446–458. https://doi.org/10.1037/edu0000296.

Alderton, D. L., J. H. Wolfe, and G. E. Larson. 1997. "The ECAT Battery." *Military Psychology* 9, no. 1: 5–37. https://doi.org/10.1207/s15327876mp0901_1.

ASVAB Enlistment Testing Program. 2020a. *What is the ASVAB?* August 13. Accessed May 18, 2022, https://www.officialasvab.com/.

ASVAB Enlistment Testing Program. 2020b. *Fairness Information*. August 13. Accessed May 18, 2022, https://www.officialasvab.com/researchers/fairness-information/.

Blankenship, T. L., M. A. Slough, S. D. Calkins, K. Deater-Deckard, J. Kim-Spoon, and M. A. Bell. 2019. "Attention and Executive Functioning in Infancy: Links to Childhood Executive Function and Reading Achievement." *Developmental Science* 22, no. 6: e12824. https://doi.org/10.1111/desc.12824.

Bosco, F., D. G. Allen, and K. Singh. 2015. "Executive Attention: An Alternative Perspective on General Mental Ability, Performance, and Subgroup Differences." *Personnel Psychology* 68, no. 4: 859–898. https://doi.org/10.1111/peps.12099.

Brown, R. D., F. L. Oswald, and P. D. Converse. 2017. "Estimating Operational Validity Under Incidental Range Restriction: Some Important but Neglected Issues." *Practical Assessment, Research, and Evaluation* 22, no. 1: 6. https://doi.org/10.7275/e7gh-0785.

Bureau of Labor Statistics, U.S. Department of Labor. 2019. *National Longitudinal Survey of Youth 1997 cohort, 1997–2017 (rounds 1–18). Produced and distributed by the Center for Human Resource Research (CHRR)*. Columbus, OH: The Ohio State University.

Burgoyne, A. P., and R. W. Engle. 2020. "Attention Control: A Cornerstone of Higher-Order Cognition." *Current Directions in Psychological Science* 29, no. 6: 624–630. https://doi.org/10.1177/0963721420969371.

Burgoyne, A. P., L. J. Harris, and D. Z. Hambrick. 2019. "Predicting Piano Skill Acquisition in Beginners: The Role of General Intelligence, Music Aptitude, and Mindset." *Intelligence* 76: 101383. https://doi.org/10.1016/j.intell.2019.101383.

Burgoyne, A. P., C. A. Mashburn, and R. W. Engle. 2021. "Reducing Adverse Impact in High-Stakes Testing." *Intelligence* 87: 101561. https://doi.org/10.1016/j.intell.2021.101561.

Burgoyne, A. P., C. A. Mashburn, J. S. Tsukahara, and R. W. Engle. 2022. "Attention Control and Process Overlap Theory: Searching for Cognitive Processes Underpinning the Positive Manifold." *Intelligence* 91: 101629. https://doi.org/10.1016/j.intell.2022.101629.

Burgoyne, A. P., J. S. Tsukahara, C. A. Mashburn, R. Pak, and R. W. Engle. 2023. "Nature and Measurement of Attention Control." *Journal of Experimental Psychology: General* 152, no. 8: 2369.

Carretta, T. R., and M. J. Ree. 2022. "Correction for Range Restriction: Lessons From 20 Research Scenarios." *Military Psychology* 34: 551–569. https://doi.org/10.1080/08995605.2021.2022067.

Conway, A. R. A., N. Cowan, M. F. Bunting, D. J. Therriault, and S. R. B. Minkoff. 2002. "A Latent Variable Analysis of Working Memory Capacity, Short-Term Memory Capacity, Processing Speed, and General Fluid Intelligence." *Intelligence* 30, no. 2: 163–183.

Cowan, N., E. M. Elliott, J. S. Saults, et al. 2005. "On the Capacity of Attention: Its Estimation and its Role in Working Memory and Cognitive Aptitudes." *Cognitive Psychology* 51, no. 1: 42–100.

Dempsey, G. M. 2015. *The National Military Strategy of the United States of America 2015*. Washington: Department of Defense-Joint Staff of Chiefs, 3–4.

Dempster, F. N. 1991. "Inhibitory Processes: A Negleted Dimension of Intelligence." *Intelligence* 15, no. 2: 157–173. https://doi.org/10.1016/0160-2896(91)90028-C.

Draheim, C., R. Pak, A. A. Draheim, and R. W. Engle. 2022. "The Role of Attention Control in Complex Real-World Tasks." *Psychonomic Bulletin & Review* 29: 1143–1197. https://doi.org/10.3758/s13423-021-02052-2.

Draheim, C., J. S. Tshukara, and R. W. Engle. 2023. "Replication and Extension of the Toolbox Approach to Measuring Attention Control." *Behavior Research Methods* 56, no. 3: 2135–2157.

Draheim, C., J. S. Tsukahara, J. D. Martin, C. A. Mashburn, and R. W. Engle. 2021. "A Toolbox Approach to Improving the Measurement of Attention Control." *Journal of Experimental Psychology: General* 150, no. 2: 242–275. https://psycnet.apa.org/doi/10.1037/xge0000783.

Engen, H. G., and M. C. Anderson. 2018. "Memory Control: A Fundamental Mechanism of Emotion Regulation." *Trends in Cognitive Sciences* 22, no. 11: 982–995. https://doi.org/10.1016/j.tics.2018.07.015.

Engle, R. W. 2002. "Working Memory Capacity As Executive Attention." *Current Directions in Psychological Science* 11, no. 1: 19–23. https://doi.org/10.1111/1467-8721.00160.

Engle, R. W. 2018. "Working Memory and Executive Attention: A Revisit." *Perspectives on Psychological Science* 13, no. 2: 190–193. https://doi.org/10.1177/1745691617720478.

Engle, R. W., S. W. Tuholski, J. E. Laughlin, and A. R. A. Conway. 1999. "Working Memory, Short-Term Memory, and General Fluid Intelligence: a Latent-Variable Approach." *Journal of Experimental Psychology: General* 128, no. 3: 309–331. https://psycnet.apa.org/doi/10.1037/0096-3445.128.3.309.

Faul, F., E. Erdfelder, A. Buchner, and A.-G. Lang. 2009. "Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses." *Behavior Research Methods* 41: 1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

Garrison, K. E., and B. J. Schmeichel. 2022. "Getting over It: Working Memory Capacity and Affective Responses to Stressful Events in Daily Life." *Emotion* 22, no. 3: 418–429. https://doi.org/10.1037/emo0000755.

Guo, Z., J. Zou, C. He, X. Tan, C. Chen, and G. Feng. 2020. "The Importance of Cognitive and Mental Factors on Prediction of Job Performance in Chinese High-Speed Railway Dispatchers." *Journal of Advanced Transportation* 2020: 1–13. https://doi.org/10.1155/2020/7153972.

Hallett, P. E. 1978. "Primary and Secondary Saccades to Goals Defined by Instructions." *Vision Research* 18: 1279–1296. https://doi.org/10.1016/0042-6989(78)90218-3.

Hambrick, D. Z., A. P. Burgoyne, E. M. Altmann, and T. J. Matteson. 2023. "Explaining the Validity of the Asvab for Job-Relevant Multitasking Performance: The Role of Placekeeping Ability." *Journal of Intelligence* 11, no. 12: 225.

Hambrick, D. Z., A. P. Burgoyne, and F. L. Oswald. 2024. "The Validity of General Cognitive Ability Predicting Job-Specific Performance Is Stable Across Different Levels of Job Experience." *Journal of Applied Psychology* 109, no. 3: 437–455. https://doi.org/10.1037/apl0001150.

Hambrick, D. Z., T. A. Rench, E. M. Poposki, et al. 2011. "The Relationship between the Asvab and Multitasking in Navy Sailors: A Process-Specific Approach." *Military Psychology* 23, no. 4: 365–380. https://doi.org/10.1080/08995605.2011.589323.

Hedge, C., G. Powell, and P. Sumner. 2018. "The Reliability Paradox: Why Robust Cognitive Tasks Do Not Produce Reliable Individual Differences." *Behavior Research Methods* 50, no. 3: 1166–1186. https://doi.org/10.3758/s13428-017-0935-1.

Held, J. D., T. R. Carretta, and M. G. Rumsey. 2014. "Evaluation of Tests of Perceptual Speed/Accuracy and Spatial Ability for Use in Military

Occupational Classification." *Military Psychology* 26, no. 3: 199–220. https://doi.org/10.1037/mil0000043.

Hough, L. M., F. L. Oswald, and R. E. Ployhart. 2001. "Determinants, Detection and Amelioration of Adverse Impact in Personnel Selection Procedures: Issues, Evidence and Lessons Learned." *International Journal of Selection and Assessment* 9, no. 1–2: 152–194. https://doi.org/10.1111/1468-2389.00171.

Hutchison, K. A. 2007. "Attentional Control and the Relatedness Proportion Effect in Semantic Priming." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 33, no. 4: 645–662. https://psycnet.apa.org/doi/10.1037/0278-7393.33.4.645.

IBM Corp. Released. 2021. *IBM SPSS Statistics for Windows, Version 28.0*. Armonk, NY: IBM Corp.

Kamarck, K. N. 2019. Diversity, Inclusion, and Equal Opportunity in the Armed Services: Background and Issues for Congress *(CRS Report R44321*. Washington, DC: Congressional Research Service.

Kane, M. J., M. K. Bleckley, A. R. A. Conway, and R. W. Engle. 2001. "A Controlled-Attention View of Working-Memory Capacity." *Journal of Experimental Psychology: General* 130, no. 2: 169–183. https://psycnet.apa.org/doi/10.1037/0096-3445.130.2.169.

Kane, M. J., D. Z. Hambrick, S. W. Tuholski, O. Wilhelm, T. W. Payne, and R. 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, no. 2: 189–217. https://psycnet.apa.org/doi/10.1037/0096-3445.133.2.189.

Landy, F. J., L. J. Shankster, and S. S. Kohler. 1994. "Personnel Selection and Placement." *Annual Review of Psychology* 45, no. 1: 261–296. https://doi.org/10.1146/annurev.ps.45.020194.001401.

Luck, S. J., and E. K. Vogel. 1997. "The Capacity of Visual Working Memory for Features and Conjunctions." *Nature* 390, no. 6657: 279–281. https://doi.org/10.1038/36846.

Martin, J., C. A. Mashburn, and R. W. Engle. 2020. "Improving the Validity of the Armed Service Vocational Aptitude Battery With Measures of Attention Control." *Journal of Applied Research in Memory and Cognition* 9, no. 3: 323–335. https://doi.org/10.1016/j.jarmac.2020.04.002.

Martin, J. D., J. S. Tsukahara, C. Draheim, et al. 2021. "The Visual Arrays Task: Visual Storage Capacity or Attention Control?" *Journal of Experimental Psychology: General* 150, no. 12: 2525–2551. https://psycnet.apa.org/doi/10.1037/xge0001048.

Mashburn, C. A., A. P. Burgoyne, and R. W. Engle. 2023. "Working Memory, Intelligence, and Life Success: Examining Relations to Academic Achievement, Job Performance, Physical Health, Mortality, and Psychological Well-Being." In *Memory in Science for Society: There is Nothing as Practical as a Good Theory*, edited by R. H. Logie, N. Cowan, S. E. Gathercole, R. W. Engle, and Z. E. Wen. Oxford University Press.

McVay, J. C., and M. J. Kane. 2012. "Why Does Working Memory Capacity Predict Variation in Reading Comprehension? On the Influence of Mind Wandering and Executive Attention." *Journal of Experimental Psychology: General* 141, no. 2: 302–320. https://doi.org/10.1037/a0025250.

Moran, T. P. 2016. "Anxiety and Working Memory Capacity: A Meta-Analysis and Narrative Review." *Psychological Bulletin* 142, no. 8: 831–864. https://doi.org/10.1037/bul0000051.

Navy Medicine. n.d. *ASTB-E Overview*. Department 53OP-Operational Psychology. Accessed November 10, 2022, https://www.med.navy.mil/Navy-Medicine-Operational-Training-Command/Naval-Aerospace-Medical-Institute/Department-53-Operational-Psychology/.

Nelson, L. C. 2003. Working My, General Intelligence, and Job Performance (Order No. 3076329). Available from ProQuest Dissertations & Theses A&I; ProQuest Dissertations & Theses Global. (305322924), https://go.openathens.net/redirector/gatech.edu?url=; https://search.proquest.com/dissertations-theses/working-memory-general-intelligence-job/docview/305322924/se-2?accountid=11107.

Nimon, K. 2010. "Regression Commonality Analysis: Demonstration of an SPSS Solution." *Multiple Linear Regression Viewpoints* 36, no. 1: 10–17.

Nimon, K. F., and F. L. Oswald. 2013. "Understanding the Results of Multiple Linear Regression: Beyond Standardized Regression Coefficients." *Organizational Research Methods* 16, no. 4: 650–674. https://doi.org/10.1177/1094428113493929.

Outtz, J. L., and D. A. Newman. 2011. "A Theory of Adverse Impact." In *Adverse Impact: Implications for Organizational Staffing and High Stakes Selection*, edited by J. L. Outtz, 53–94. Routledge.

Raven, J., and J. H. Court (1998). Manual for Raven's Progressive Matrices and Vocabulary Scales.

Redick, T. S., R. P. Heitz, and R. W. Engle. 2007. "Working Memory Capacity and Inhibition: Cognitive and Social Consequences." In *Inhibition in Cognition*, edited by D. S. Gorfein, and C. M. MacLeod, 125–142. American Psychological Association. https://doi.org/10.1037/11587-007.

Roberts, R. D., G. N. Goff, F. Anjoul, P. C. Kyllonen, G. Pallier, and L. Stankov. 2000. "The Armed Services Vocational Aptitude Battery (ASVAB)." *Learning and Individual Differences* 12, no. 1: 81–103. https://doi.org/10.1016/S1041-6080(00)00035-2.

Sackett, P. R., and H. Yang. 2000. "Correction for Range Restriction: An Expanded Typology." *Journal of Applied Psychology* 85: 112–118.

Sackett, P. R., C. Zhang, C. M. Berry, and F. Lievens. 2022. "Revisiting Meta-Analytic Estimates of Validity in Personnel Selection: Addressing Systematic Overcorrection for Restriction of Range." *Journal of Applied Psychology* 107, no. 11: 2040–2068.

Salthouse, T. A., T. M. Atkinson, and D. E. Berish. 2003. "Executive Functioning as a Potential Mediator of Age-Related Cognitive Decline in Normal Adults." *Journal of Experimental Psychology: General* 132, no. 4: 566–594. https://psycnet.apa.org/doi/10.1037/0096-3445.132.4.566.

Schmidt, F. L., W. L. Dunn, and J. E. Hunter. 1995. *Potential Utility Increases From Adding New Tests to the Armed Services Vocational Aptitude Battery (ASVAB)*. Triangle Park, NC: Battelle Memorial Inst Research.

Schmidt, F. L., J. E. Hunter, and V. W. Urry. 1976. "Statistical Power in Criterion-Related Validation Studies." *Journal of Applied Psychology* 61, no. 4: 473–485. https://psycnet.apa.org/doi/10.1037/0021-9010.61.4.473.

Shipstead, Z., D. R. B. Lindsey, R. L. Marshall, and R. W. Engle. 2014. "The Mechanisms of Working Memory Capacity: Primary Memory, Secondary Memory, and Attention Control." *Journal of Memory and Language* 72: 116–141. https://doi.org/10.1016/j.jml.2014.01.004.

Thorndike, R. L. 1949. *Personnel Selection*. New York: Wiley.

Titchener, E. B. 1908. *Lectures on the Elementary Psychology of Feeling and Attention*. New York.

Tsukahara, J. S., and R. W. Engle. 2023. *Sustaining the Focus of Attention and How it Relates to Performance in Complex Cognitive Tasks*. OSF. https://osf.io/wd5kz/download.

Tsukahara, J. S., T. L. Harrison, C. Draheim, J. D. Martin, and R. W. Engle. 2020. "Attention Control: The Missing Link Between Sensory Discrimination and Intelligence." *Attention, Perception, & Psychophysics* 82, no. 7: 3445–3478. https://doi.org/10.3758/s13414-020-02044-9.

Uniform Guidelines on Employee Selection Procedures. (1978). *29 CFR § 1607*.

Unsworth, N., B. D. McMillan, G. A. Brewer, and G. J. Spillers. 2012. "Everyday Attention Failures: an Individual Differences Investigation."

Journal of Experimental Ppsychology. Learning, Memory, and Cognition 38, no. 6: 1765–1772. https://psycnet.apa.org/doi/10.1037/a0028075.

Unsworth, N., J. C. Schrock, and R. W. Engle. 2004. "Working Memory Capacity and the Antisaccade Task: Individual Differences in Voluntary Saccade Control." *Journal of Experimental Psychology. Learning, Memory, and Cognition* 30, no. 6: 1302–1321.

Valentine, Jr. L. D. 1977. *Prediction of Air Force Technical Training Success From ASVAB and Educational Background*. Texas: AIR Force Human Resources Lab Brooks AFB.

Wegner, T. G., and M. J. Ree. 1986. *Alternative Armed Forces Qualification Test Composites*.Texas: Air Force Human Resources Lab Brooks AFB.

Welsh, Jr. J. R., S. K. Kucinkas, and L. T. Curran. 1990. *Armed Services Vocational Battery (ASVAB): Integrative Review of Validity Studies*. Operational Technologies Corp. San Antonio, TX.

Westergard, C. 2019. "You Catch More Flies With Honey: Reevaluating the Erroneous Premises of the Military Exception to reTitle VII." *Marquette Benefits and Social Welfare Law Review* 20, no. 2: Article 5. https://scholarship.law.marquette.edu/benefits/vol20/iss2/5.

Wigdor, A. K., and B. Green, eds. 1991. *Performance Assessment in the Workplace*, vols. I & II. National Academy Press.

William Revelle 2024. Psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.4.6, https://CRAN.R-project.org/package=psych.

Wise, L., J. Welsh, F. Grafton, et al. 1992. Sensitivity and Fairness of the Armed Services Vocational Aptitude Battery (ASVAB) Technical Composites. Defense Manpower Data Center, Department of Defense: Personnel Testing Division. https://www.officialasvab.com/wp-content/uploads/2019/08/AS92009_Sensitivity_Fairness_of_ASVAB_Tech_Composites.pdf.

Wise, L. L. 1994. "Setting Performance Goals for the DoD Linkage Model." In *Modeling Cost and Performance for Military*, Edited by B. F. Green, and A. S. Mavor. National Academies Press. https://doi.org/10.17226/2344.

Woltz, D. J. 1988. "An Investigation of the Role of Working Memory in Procedural Skill Acquisition." *Journal of Experimental Psychology: General* 117, no. 3: 319–331. https://psycnet.apa.org/doi/10.1037/0096-3445.117.3.319.

Zedeck, S. 2011. "Adverse Impact: History and Evolution." In *Adverse impact: Implications for Organizational Staffing and High Stakes Selection*, edited by J. L. Outtz, 3–28. Routledge.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.