Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 2077–2088
https://doi.org/10.54941/ahfe1008434

Continued Assessment of the Shortened Attention Control Tasks: Impact of Practice Effects, an Interfering Pre-cue, and Speed-Accuracy Manipulation

Christopher Draheim[1], Ciara Sibley[1], Kaylin Strong[1], Michael Kukenberger[2,3], and Joseph T. Coyne[1]

[1]Naval Research Laboratory, Washington, DC 20375, USA
[2]Naval Aerospace Medical Institute, Pensacola, FL 32508, USA
[3]University of New Hampshire, Durham, NH 03824, USA

ABSTRACT

Our previous research has demonstrated the viability of three-minute long ("squared") double conflict attention control tests for predicting early flight training outcomes in U.S. Naval flight students. The present study was an effort to continue exploring and refining these tests by assessing both learning and the impact of two experimental manipulations (an interfering pre-cue and a speed-accuracy manipulation) in the Double Stroop task, a shortened variant of the traditional color Stroop task. We administered 12 Double Stroop tasks, two Double Flanker tasks, three processing speed tasks, and a survey assessing video gaming experience to a final sample of 131 Naval flight students and 64 enlisted Sailors ($N = 195$). We found the pre-cue interval and order of administration had a substantive impact on performance, as did the speed-accuracy manipulation. Further, learning seemed to plateau after seven attempts of the Double Stroop task (i.e., after around 10 minutes). Processing speed was moderately related to performance on the double conflict tasks. Action video gaming engagement was mostly unrelated to improvements, although there was weak evidence that gamers showed slightly larger gains with practice. These results may help understand the nature of performance and processes at play in these double conflict tests more generally.

Keywords: Attention control, Individual differences, Personnel selection

INTRODUCTION

Successfully identifying people capable of becoming a pilot has been a persistent challenge for militaries since around the time of World War I (Griffin & Koonce, 1996). To aid in this effort, thousands of applicants across multiple branches of the U.S. military take the Aviation Selection Test Battery (ASTB) each year. The U.S. Navy uses it to assess the cognitive ability of the examinee to determine whether they have the potential to become a Naval Aviator (pilot) or Flight Officer (navigator). The ASTB consists of tests assessing crystalized knowledge (e.g., math, verbal, and aviation knowledge), fluid abilities (e.g., spatial ability and attention), and psychomotor ability.

Received October 16, 2026; Revised November 25, 2026; Accepted December 7, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

As such, the ASTB combines achievement, aptitude, and motor-based assessments into one comprehensive battery.

One of the challenges of the ASTB is that it is becoming increasingly difficult in the modern internet era to ensure potential test-takers are naïve to the content of the battery. That is, to maintain validity, cognitive selection tests typically require the examinee to have had minimal previous experience with it, as well as minimal prior knowledge of the specific test content or potential strategies to improve performance. But with a high-stakes test in which performance has a direct impact on the examinee's life *and* in a highly competitive area such as military aviation with many more applicants than spots, test security is severely threatened due to having highly motivated candidates coupled with easy access to online test preparation materials and previous examinees. To that end, we have observed a significant reduction in the reliability and predictive validity of some ASTB subtests over time (e.g., Coyne et al., 2022).

ATTENTION CONTROL

Due to the challenges facing the ASTB, the U.S. Naval Research Laboratory has engaged in research to develop and iterate on cognitive tests which may be both more predictive of Navy flight school outcomes while better withstanding the rigors of selection testing among highly motivated candidates. A promising construct for this purpose is ***attention control.*** Attention control is the ability to maintain goal-directed behavior and goal-relevant information while ignoring, suppressing, or disengaging from goal-irrelevant information and behaviors. It is particularly important in cognitively demanding situations, such as those involving novel situations, high cognitive load, external distraction from the environment, and/or internal distraction from one's internal thoughts and cognitive processes.

It has been argued that attention control is the driver of higher-order cognition in humans and possibly the best marker of an individual's cognitive *potential* (e.g., Draheim et al., 2022). This view is supported by a growing body of research demonstrating the importance of attention control in real-world human performance and behavior (e.g., Draheim et al., 2022) and that when comprehensively measured, attention control explains the relationship between fluid intelligence and working memory capacity (Draheim et al., 2021). It is therefore not surprising that modern attention control tests have been found to provide incremental predictive validity to existing Navy selection batteries (e.g., Burgoyne et al., 2025).

The Double Conflict Tasks

A class of attention control tests that has emerged from recent research are double conflict or "squared" variants of the traditional cognitive conflict tests of inhibition/attention control that were specifically designed and validated for individual differences purposes (see Burgoyne et al., 2023). These double conflict tasks are labeled as such because interference (i.e., cognitive conflict) can occur in either or both the stimulus and the response options. They are especially appealing for use in personnel selection thanks to their improved

psychometrics from traditional conflict tasks (see Draheim et al., 2021), relative ease of administration, and very quick administration time. Further, potential concerns over their process purity (i.e., how purely performance on them reflects a single underlying mechanism/ability) is less important when the goal is to identify promising candidates and predict future incomes, as opposed to isolating performance variance to a single underlying construct. Fig. 1 illustrates our version of the Double Stroop Task and Fig. 2 illustrates our version of the Double Flanker Task, the two relevant for the present study.

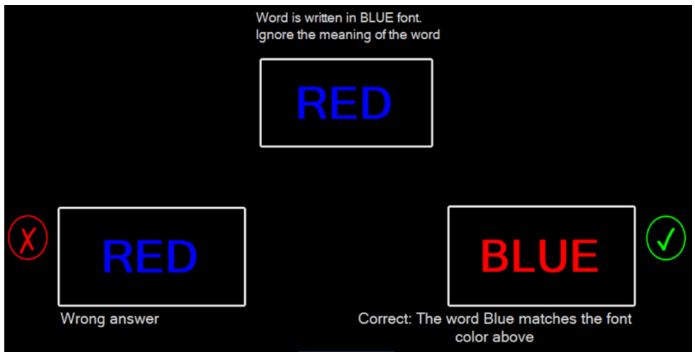


Figure 1: The double stroop task. Participants are shown three color words ("RED" or "BLUE") in either red or blue font: one near the center of the screen (stimulus) with two response options in boxes underneath. The stimulus and response options can be congruent (the font color matches the word; e.g., "RED") or incongruent (the font color is different from the word; e.g., "RED"). The goal is to select the response option with the same semantic meaning as the font color of the stimulus above it. After instructions and 30 seconds of practice with feedback, participants have 90 seconds to respond to as many trials as possible, with a 200 ms intertrial interval. The score is calculated by subtracting the number of incorrect responses from the number of correct responses. The above image is taken from the instructions phase.

Previous research established the reliability and validity of these double conflict tests of attention control and that they can predict training outcomes from Naval Flight School (e.g., Coyne et al., 2025). However, efforts are ongoing to improve and iterate upon them should they eventually be used for aviation-related selection. For example, one area of improvement is that we have consistently found that the double Flanker to be more predictive of training outcomes than the double Stroop. We also have hypothesized a number of potential manipulations that may magnify individual differences (and therefore discriminability) while preventing, or at least reducing, candidates' ability to engage in some undesirable strategies to "game" the test such that performance is more indicative of their preparation than their true ability. Finally, should these double conflict tests be used for personnel selection, it needs to be assessed how well they hold up across repeated attempts from highly motivated prospective candidates. That is, does extensive practice on them prior to selection potentially undermine their psychometric properties and/or predictive validity? These were the questions we sought to explore in the present protocol by administering a series of double conflict attention control tests along with some other measures of interest.

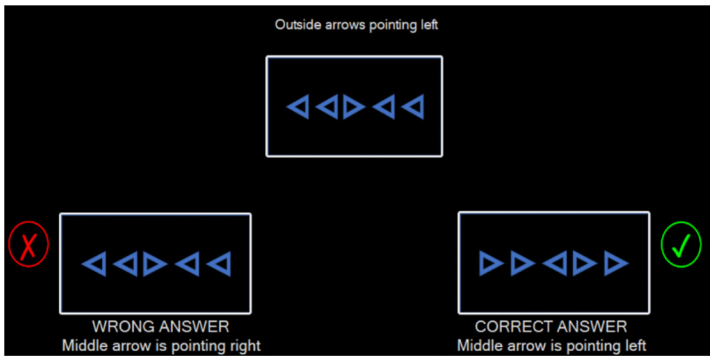


Figure 2: The double flanker task. Participants are shown three groups of arrows pointing either left or right: one near the top of the screen (stimulus) with two response options underneath. Each arrow group can be congruent (all arrows pointing in the same direction; e.g., ← ← ← ← ←) or incongruent (the central arrow pointing in a different direction from the flanking arrows; e.g., ← ← ← ← ←). The goal is to select the response option with its central arrow pointing in the same direction as the flanking arrows of the target stimulus. After instructions and 30 seconds of practice with feedback, participants are given 90 seconds to respond to as many trials as possible, with a 200 ms intertrial interval. The score is calculated by subtracting the number of incorrect responses from the number of correct responses. The above image is taken from the instruction phase, and the top box can appear centrally or on the left- or right-part of the screen.

METHOD

Participants

A total of 216 Naval flight students (NFS) and enlisted Sailors consented to participate. NFS are officers who have been accepted into Naval Flight School for training to become a Naval Aviator or Flight Officer, Enlisted Sailors were from a variety of ratings. Participants were volunteers and thus not directly compensated but were aware that their participation may help improve personnel selection within the Navy. The study was approved by the Naval Research Laboratory's Institutional Review Board (Protocol No. NRL.2021.0003).

Procedure

Participants performed a series of 14 double conflict attention control tests followed by three processing speed tests within a single session lasting approximately one hour. Eye tracking and pupillometry data were collected but are not discussed here. Various surveys were interspersed throughout the protocol, only a subset of the video gaming survey is analyzed here.

Double Conflict Attention Control Tasks

1. *Double Flanker*. This task is identical to the one shown in Fig. 2.
2. *Double Stroop – Standard*. This task is identical to the one in Fig. 1.

3. *Double Stroop – Pre-Cue*. This task is identical to the one shown in Fig. 1, except prior to the response options appearing, the target stimulus briefly appeared in a neutral color (white). This manipulation was used by Liao et al. (2025) in a traditional color Stroop task to increase cognitive interference by presenting the to-be-ignored element prior to the trial. Participants performed five versions of this task, each with different pre-cue intervals (100, 150, 200, 250, or 300 ms). Based on whether their participant ID was odd or even, they performed these five pre-cue tasks either in ascending or descending order of pre-cue interval.
4. *Double Stroop – Speed-Accuracy*. For NFS, these tasks included a response deadline in which failure to respond in time would result in the trial being scored as incorrect (1500 ms for the first task; 1250 ms for the second). For enlisted Sailors, there was no response deadline but instead the instructions urged them to prioritize accuracy over speed.

Processing Speed

The three processing speed tests were simple clerical tests administered from The Psychological Experiment Building Language (PEBL) Test Battery (Mueller & Piper, 2014). Specifically, these were the Letter-Digit Substitution Task (60 trials), Four-Choice Response Task (80 trials), and Pattern Comparison Task (80 trials).

Video Game Survey

The video gaming survey was a modified version of the Bavelier Lab Video Game Questionnaire (Green et al., 2017). Among other questions, the survey asked participants to report "...the number of hours played per week when you played the most" across nine different gaming genres. There were six response options for each question: "Never", "Less than one hour", "Between 1 and 3 hours", "Between 3 and 5 hours", "Between 5 and 10 hours", and "More than 10 hours", which we coded as 0–5, respectively, for data analysis. The survey asked participants to separately report both on their childhood (aged 3–13 years) and current (within the last 12 months) levels of engagement.

RESULTS

Data Processing

Non-positive double conflict task scores were set to missing. Then, any participant with more than 4 missing Double Stroop scores were removed from analysis. This resulted in data being retained for $N = 195$ (131 NFS; 64 enlisted).

For processing speed, we entered correct trial-only mean RT from each of the three tasks into an exploratory factor analysis and obtained regression-based factor scores to get a single score for performance.

For action video gaming experience, we summed responses from four categories on the gaming survey (First/Third-Person Shooters, Action-RPG/Adventure, Sports/Driving, and Real-Time Strategy/MOBA) to create a score

for action video gaming experience, with separate scores for childhood vs. current (i.e., adult) gaming engagement. As responses from each question were coded as 0–5, the aggregated action gaming engagement score could range from 0 to 20, again with separate scores for childhood and adult engagement.

Descriptives, Reliability, and Gains

Descriptive statistics for each variable of interest are presented in Table 1. And aggregated descriptives for the double conflict tasks are presented in Table 2.

We assessed internal consistency for each task via even-odd split-half corrected with the Spearman-Brown Prophecy formula. Internal consistency was high for all double conflict tasks, ranging from .89 to .94. Test-retest reliabilities for consecutive tasks of the same type were also generally strong: for the Double Stroop – Standard tasks it ranged from $r = .74$ - .80; for the Double Stroop – Pre-cue tasks it ranged from $r = .72$ - .83, and for the Double Stroop – Speed-Accuracy tasks it was $r = .87$ for enlisted Sailors and $r = .66$ for NFS. Test-retest for the Double Flanker task was $r = .74$.

Table 1: Descriptive statistics for all tasks.

Task	Enlisted			NFS		
	n	AT	SD	*n*	AT	SD
Flanker 1	58	13.7	12.5	130	23.6	9.8
Stroop Standard 1	59	23.3	14.5	128	31.7	12.3
Stroop Pre-cue 1	62	30.7	17.9	131	41 3	14.7
Stroop Pre-cue 2	63	35.9	16.0	131	45.0	12.2
Stroop Pre-cue 3	64	37.6	15.4	130	47.7	11.4
Stroop Pre-cue 4	64	40.1	15.3	130	49.6	12.0
Stroop Pre-cue 5	62	44.2	14.4	131	50.6	12.5
Stroop Standard 2	63	45.7	13.9	131	52.9	11.2
Stroop Standard 3	64	44.5	14.7	131	51.3	12.0
Stroop Standard 4	64	43.6	13.3	129	51.5	10.4
Stroop SAT 1	64	45.3	14.1	125	33.9	19.8
Stroop SAT 2	64	46.7	13.1	106	31.7	18.8
Stroop Standard 5	61	46.3	11.8	126	50.9	12.3
Flanker 2	52	30.4	11.9	128	33.2	8.6
Four-Choice	57	549.4	166.3	124	473.9	89.6
Letter-Digit Sub.	56	2055.3	425.9	126	1843.7	391.6
Pattern Comparison	54	2012.8	563.7	121	1644.1	397.5
Adult Action Gaming	63	6.2	5.1	124	4.4	4.2

(Continued)

Table 1: Continued.

Task	Enlisted			NFS		
	n	AT	SD	*n*	AT	SD
Child Action Gaming	61	7.3	4.9	124	7.0	4.8
Adult Flight Sim.	63	1.1	1.7	124	0.8	1.3
Child Flight Sim.	61	0.9	1.6	124	1.0	1.4

Note. Tasks are shown in the order they were administered except for the gaming survey. SAT = Speed-accuracy tradeoff; Sub = substitution; Sim = simulation.

Table 2: Aggregated double conflict task descriptives.

Task Type	Enlisted			NFS		
	n	*M*	*SD*	*n*	*M*	*SD*
Stroop Standard	55	41.5	12.4	123	48.0	10.0
Stroop Pre-Cue	60	38.6	14.3	129	47.2	10.9
Stroop SAT	64	46.0	13.2	105	36.3	17.3
Flanker	52	24.2	11.3	128	33.3	8.6

Improvements on each Double Stroop – Standard task are shown in Table 3. Improvement on the second administration of the Double Flanker task was $d = 0.94$ for enlisted Sailors and $d = 0.98$ for NFS.

Table 3: Improvements on the double stroop – standard tasks.

Task	Enlisted			NFS		
	Score	*d*	Total *d*	Score	*d*	Total *d*
Stroop Standard 1 [1]	23.8	—	—	31.7	—	—
Stroop Standard 2 [7]	45.7	1.50	1.50	52.9	1.65	1.65
Stroop Standard 3 [8]	44.5	−0.08	1.42	51.3	−0.13	1.41
Stroop Standard 4 [9]	43.6	0.11	1.47	515	0.02	1.43
Stroop Standard 5 [12]	46.3	0.19	1.55	50.9	−0.05	1.33

Note. Numbers in brackets are the order of the task compared to all other Stroop tasks, e g., Stroop Standard 5 was the 12th and final Stroop task administered in this study, if is the Cohen's *d* improvement from the previous Stroop Standard: Total *d* is the improvement from Stroop Standard 1. The SD from Stroop Standard 1 was used for all ds.

Impact of the Double Stroop Pre-Cue Interval

Table 4 shows performance broken down by pre-cue interval and counterbalance condition (ascending order vs. descending order of pre-cue). Figure 3 shows performance of the two conditions by both order and pre-cue interval.

Table 4: Descriptives based on double stroop pre-cue interval (all participants).

Pre-Cue	Ascender		Descender		Combined	
Interval (ms)	*M*	*SD*	*M*	*SD*	*M*	*SD*
100	34.3	16.0	48.0	12.5	41.4	15.8
150	39.S	14.0	46.4	13.5	43.2	14.1
200	43.0	14 0	45.1	13.4	44.4	13.7
250	46.5	14.7	44.2	14.0	45.3	14.4
JOO	49.1	14 4	41.3	16.3	45.1	15.9
Combined	43.4	13.0	45.5	12.4	44.5	12.7

Note. Ascenders performed the tasks in ascending order (i.e, 100 ms first); Descenders performed the tasks in descending order (i.e, 300 ms first). Ascender *n* ranged from 93–94; Descender *n* ranged from 99–101.

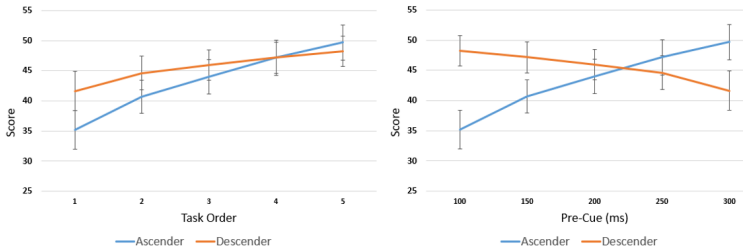


Figure 3: Performance on Pre-Cue Tasks Based on Counterbalance Condition (All Participants. The left panel shows performance based on temporal order participants performed the task. The right panel shows these same data but based on pre-cue time of the task (i.e., the descender line is flipped). Error bars are the 95% CI.

We conducted a 5 × 2 mixed ANOVA assessing performance of the pre-cue tasks based on the order they were administered (within-subjects) and counterbalance condition (ascending vs. descending order; between-subjects; e.g., the left-panel of Fig. 3). This model revealed a significant within-subject main effect of task, $F(4, 748) = 80.255$, $p < .001$, $\eta^2_p = .30$, but not for the between-subject factor of counterbalance condition, $F(1, 187) = 1.389$, $p = .240$. This indicated that participants improved on the pre-cue tasks across administrations regardless of which order they performed them. We also found a significant interaction between order and counterbalance condition, $F(4, 748) = 11.600$, $p < .001$, $\eta^2_p = .06$. Bonferroni-adjusted pairwise comparison tests indicated ascenders performed worse than descenders on both their first and second task ($M_{dif} = 6.42$; $p = .006$ and $M_{dif} = 3.93$; $p = 047$, respectively). These pairwise comparisons also revealed that overall performance improved from each pre-cue task to the next, however this was driven by the ascenders, as the descenders only showed significant improvement from their first to second task.

To assess the impact of the specific pre-cue intervals, we conducted another 5 × 2 repeated measures ANOVA but the within-subjects variable was entered based on their pre-cue interval instead of the order

the task was administered (e.g., the right panel of Fig. 3). This model indicated a significant within-subjects main effect of pre-cue interval, $F(4, 748) = 13.627$, $p < .001$, $\eta^2_p = .07$, and a significant interaction between pre-cue interval and counterbalance condition, $F(4, 748) = 78.228$, $p < .001$, $\eta^2_p = .30$. Bonferroni-adjusted pairwise comparisons revealed that performance on the 100 ms pre-cue task was worse than each of the other pre-cue tasks (M_{dif} ranging from 2.20 – 4.16; all ps < .004), and that performance on the 150 ms pre-cue was significantly worse than performance on the 250 ms pre-cue ($M_{dif} = 1.96$, $p = .014$). These tests also revealed that ascenders performed worse than descenders on both the 100 ms and 150 ms pre-cues ($M_{dif} = 13.02$, $p < .001$ and $M_{dif} = 6.47$, $p < .001$, respectively) but ascenders performed better on the 300 ms pre-cue task ($M_{dif} = 8.10$, $p < .001$).

Association Between Performance and Other Variables

We computed Pearson correlations to assess the relationship between the processing speed factor score and double conflict performance. Processing speed had a fairly strong correlation to performance on the double conflict tasks (*r*s ranging from .30 to .45; excluding the two Double Stroop – Speed-Accuracy task). However, this was driven by the enlisted Sailors with an *n* of only around 50. We therefore assessed correlations among NFS only and found more modest correlations ranging from $r = .13$ to .39 between processing speed and double conflict task performance. Processing speed's strongest correlations were unsurprisingly with the two response deadline Double Stroop tasks ($r = .31$ and .39, $p < .001$); the weakest was with the first Double Stroop ($r = .13$, *ns*).

We computed Pearson correlations to assess the relationship between video gaming experience and the double conflict task scores, specifically for NFS. Correlations were generally non-significant with some exceptions: adult action gaming engagement significantly correlated with performance on Double Stroop – Standard 2 ($r = .23$), Standard 3 ($r = .19$), and both Double Stroop tasks with a response deadline ($r = .30$ and .23, respectively). Childhood action gaming engagement was significantly correlated with Double Stroop – Standard 2 ($r = .20$), Double Stroop – Standard 3 ($r = .19$), and the first Double Stroop with a response deadline ($r = .21$). We calculated a difference score for improvement from the first to second Double Stroop – Standard task and found it was significantly correlated with childhood action gaming levels ($r = .23$).

DISCUSSION

Improvement on the Double Conflict Tasks

Analysis of Cohen's *d* improvement in the Double Stroop tasks revealed that participants continued to improve until the eighth task, i.e., just over 10 minutes of performing the Double Stroop, at which point participants had improved by around $d = 1.50$. Improvement on the second Double Flanker

task from the first was just under $d = 1.00$. It is interesting that enlisted Sailors and NFS showed fairly similar Cohen's d improvements on both types of double conflict tasks.

Pre-Cue Interval

We failed to find a significant difference in average performance on the five pre-cue tasks based on the order they were administered. Further, participants improved on each task regardless of the pre-cue interval order. However, both pre-cue interval and order did impact performance. Analyses revealed that the 100 ms pre-cue produced lower scores, and also that participants who received the 100 ms pre-cue task first improved more on subsequent tasks (as the pre-cue increased). This may suggest a carryover effect and/or that the ordering of this interference-inducing manipulation affects learning. Further research is warranted to better explore this, ideally with the pre-cue interval manipulated on a trial-by-trial (instead of task-level) basis, keeping the intertrial interval consistent regardless of pre-cue time, and possibly increasing the intertrial interval to reduce its possible interaction with pre-cue interval. Nonetheless, this manipulation does seem to show promise if the goal is to increase cognitive interference and/or difficulty in these sorts of conflict tasks.

Impact of the Speed-Accuracy Manipulations

As expected, the response deadline manipulation in the Double Stroop had a strong, negative, effect on NFS performance, with the quicker (1250 ms) deadline resulting in substantially worse performance than the 1500 ms deadline.

The simple instruction for enlisted Sailors to prioritize accuracy over speed in the two Double Stroop – Speed-Accuracy tasks resulted in a numerical (but not statistically significant) improvement on performance over the previous Double Stroop task. It is interesting that when comparing Double Stroop 9 vs. 12 performance for enlisted vs. NFS (i.e., the tasks bookending the ones with the speed-accuracy manipulation), enlisted Sailors slightly improved from task 9 to 12 after receiving instruction to prioritize accuracy, whereas NFS performed slightly worse on task 12 relative to 9 after performing tasks requiring them to speed up. It is worth exploring in future studies how optimally participants are balancing speed and accuracy in these tasks and whether incentivizing accuracy over speed may boost performance in certain populations.

Association With Processing Speed and Video Gaming

Processing speed was generally, albeit weakly, correlated with double conflict task performance. The numerically stronger correlation with processing speed and the response deadline tasks among NFS was also expected. It is noteworthy that the weakest correlation with processing speed was that of the first Double Stroop task ($r = .13$, *ns*), perhaps indicating that when first learning the task one's processing speed is not an important determinant of performance.

Both self-reported adult and childhood engagement in action video gaming was generally unrelated to performance on the double conflict tasks. There were some exceptions, notably with both current and childhood action video gaming correlating modestly with performance on some of the Double Stroop tasks in the middle of the protocol. We also found childhood action gaming engagement was significantly correlated with improvement from the first to second Double Stroop – Standard tasks. As such, there is some hint that video gaming could confer a benefit to performance on the task during the learning process. This idea is consistent with some previous research reporting that childhood action gaming in particular may facilitate learning in cognitive tasks (e.g., Draheim et al., 2025; Zhang et al., 2021), albeit with the caveat that effects sizes are small and that action gaming levels are assessed via self-report.

REFERENCES

Burgoyne, A.P., Mashburn, C.A., Tsukahara, J.S., Pak, R., Coyne, J.T., Foroughi, C., Sibley, C., Drollinger, S.M. and Engle, R.W. (2025), Attention control measures improve the prediction of performance in Navy trainees. *International Journal of Selection and Assessment, 33*: e12510. https://doi.org/10.1111/ijsa.12510

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). Nature and measurement of attention control. *Journal of Experimental Psychology. General, 152(*8), 2369–2402. https://doi.org/10.1037/xge0001408

Coyne, J. T., Dollinger, S., Brown, N., Foroughi, C., Sibley, C., & Phillips, H. (2022). Limitations of current spatial ability testing for military aviators. *Military Psychology, 34*(1), 33–46. https://doi.org/10.1080/08995605.2021.1965786

Coyne, J. T., Jamison, L., Strong, K., Sibley, C., Foroughi, C., & Melick, S. (2025). Process-based measures in high-stakes testing: Practical implications for construct validity within military aviation selection. *Cognitive Research: Principles and Implications, 10*(1), 51. https://doi.org/10.1186/s41235-025-00660-3

Draheim, C., Sibley, C., Herdener, N., Cochrane, A., Melick, S. R., Strong, K., & Coyne, J. T. (2025). High-stakes psychomotor ability assessment: A military selection case study of practice effects in airplane tracking tasks. *Cognitive Research: Principles and Implications, 10*(1), 82. https://doi.org/10.1186/s41235-025-00672-z

Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review, 29*(11), 1143–1197. https://doi.org/10.3758/s13423-021-02052-2

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

Green, C. S., Kattner, F., Eichenbaum, A., Bediou, B., Adams, D. M., Mayer, R. E., & Bavelier, D. (2017). Playing some video games but not others is related to cognitives abilities: A critique of Unsworth et al. (2015). *Psychological Science, 28*(5), 679–682. https://doi.org/10.1177/0956797616644837

Griffin, G. R., & Koonce, J. M. (1996). Review of psychomotor skills in pilot selection research of the U.S. military services. *The International Journal of Aviation Psychology, 6*(2), 125–147. https://doi.org/10.1207/s15327108ijap0602_2

Liao, X., Sun, Y., Tian, Z., Li, Y., Fan, Y., & Li, X. (2026). The impact of interference rate on the test-retest reliability of the Stroop test. *Quarterly Journal of Experimental Psychology (2006)*, 17470218261426913. Advance online publication. https://doi.org/10.1177/17470218261426913

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *Journal of Neuroscience Methods*, *222*, 250–259. https://doi.org/10.1016/j.jneumeth.2013.10.024

Zhang, R. Y., Chopin, A., Shibata, K., Lu, Z. L., Jaeggi, S. M., Buschkuehl, M., Green, C. S., & Bavelier, D. (2021). Action video game play facilitates "learning to learn." *Communications Biology, 4*(1), 1154. https://doi.org/10.1038/s42003-021-02652-7