Replication and extension of the toolbox approach to measuring attention control

Christopher Draheim[1] · Jason S. Tshukara[2] · Randall W. Engle[2]

Abstract

There is an increasing consensus among researchers that traditional attention tasks do not validly index the attentional mechanisms that they are often used to assess. We recently tested and validated several existing, modified, and new tasks and found that accuracy-based and adaptive tasks were more reliable and valid measures of attention control than traditional ones, which typically rely on speeded responding and/or contrast comparisons in the form of difference scores (Draheim et al. *Journal of Experimental Psychology: General, 150*(2), 242–275, 2021). With these improved measures, we found that attention control fully mediated the working memory capacity–fluid intelligence relationship, a novel finding that we argued has significant theoretical implications. The present study was both a follow-up and extension to this "toolbox approach" to measuring attention control. Here, we tested updated versions of several attention control tasks in a new dataset ($N = 301$) and found, with one exception, that these tasks remain strong indicators of attention control. The present study also replicated two important findings: (1) that attention control accounted for nearly all the variance in the relationship between working memory capacity and fluid intelligence, and (2) that the strong association found between attention control and other cognitive measures is not because the attention control tasks place strong demands on processing speed. These findings show that attention control can be measured as a reliable and valid individual differences construct, and that attention control shares substantial variance with other executive functions.

Keywords Individual differences · Attention control · Measurement · Reliability paradox

The ability to manage severely limited cognitive resources is a highly important aspect of human cognition (e.g., Norman & Shallice, 1986; Turner and Engle, 1989). Specifically, our research group has argued that *attention control* is the primary driver of higher-order cognition and therefore underlies individual differences in a wide range of cognitive behaviors (e.g., Draheim et al., 2022; Burgoyne & Engle, 2020; Engle, 2002; Shipstead et al., 2016; Tsukahara et al., 2020; also see Conway et al., 2021 and Rueda, 2018, for similar perspectives). Broadly speaking, we define attention control as the ability to manage goal-directed action through a combination of maintaining relevant information and behavior and, blocking, filtering, or otherwise suppressing irrelevant information and inappropriate thought and behavior.

The nature and assessment of individual differences in attention control has become a hot area of research in recent years as researchers continue to debate the conceptualization, importance, and measurement of the construct (see von Bastian et al., 2020 for a review). A central issue to this debate is which psychological tasks are best suited to assess individual differences in the mechanisms that comprise the broader concept of attention control. Individual differences in attention control and related concepts, such as inhibitory control and cognitive control, are commonly assessed using classic psychology tasks originally employed in experimental work, such as the ubiquitous Stroop and flanker tasks. However, researchers are increasingly recognizing the shortcomings of existing paradigms, *especially when used to assess individual differences* (e.g., Draheim et al., 2016, 2019; Hedge et al., 2018; Paap & Sawi, 2016; Rouder & Haaf, 2019; Rouder et al., 2019). This trend that popular experimental tasks are often poorly suited to individual differences

✉ Christopher Draheim
cdraheim90@gmail.com

1 Department of Psychology, Lawrence University, Appleton, WI, USA

2 School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA

research was labeled the reliability paradox by Hedge et al. (2018). The reliability paradox is particularly salient within the assessment of attention or inhibition as it has consistently been shown that many highly regarded attention-based experimental tasks share virtually no meaningful variance with one another. The Stroop and flanker tasks, for example, typically share less than 2% of their performance variance despite both being considered quintessential inhibitory tasks that require respondents to resolve cognitive conflict (e.g., Rouder & Haaf, 2019). It is also important to recognize that the reliability paradox has implications for the measurement of many other cognitive abilities (e.g., task-switching and language ability) and is applicable to many areas of study within the psychological and behavioral sciences more broadly (see Draheim et al., 2019 for a review).

While the existence of the reliability paradox is not particularly controversial, there is strong debate as to why popular attention control tasks exhibit poor validity, the theoretical implications of this, and potential solutions to it (e.g., Draheim et al., 2021; Hedge et al., 2021; Rey-Mermet et al., 2018, 2019; Rouder et al., 2019, 2022). The reason for the reliability paradox is likely multifaceted, and we have emphasized two important limitations of many tasks used in differential research, (1) reliance on difference scores (i.e., the subtraction method), which attenuates reliability thereby also reducing validity, and (2) failure to properly control or account for differences and interactions between speed and accuracy (see Draheim et al., 2019, 2021). The result is that scores based on tasks such as cognitive conflict tasks (e.g., Stroop, flanker, Simon) generally have low reliability, poor validity, and are not process-pure because scores are contaminated with variance attributable to either processing speed and/or speed–accuracy tendencies and interactions. That is, participants may score better or worse on the task depending on their level of processing speed or how they balance speed vs. accuracy in these tasks, independent of their ability to control their attention. Supporting this view, Hedge et al. (2021) applied advanced modeling to several datasets and found that the small amount of variance shared across the cognitive conflict tasks was due to response cautiousness and processing speed, and not because the tasks required inhibition or resolving cognitive conflict. In other words, it appears that tasks such as Stroop and flanker do not measure the cognitive processes that they are widely believed to measure.

The toolbox approach to measuring attention control

One of the primary goals of a recent study in our lab was to develop and validate tasks specifically designed for assessing individual differences in attention control (Draheim et al., 2021). Our hypothesis was that psychometrically stronger tasks could be created by minimizing the demand for speeded responding, avoiding difference scores or other contrast comparisons, and emphasizing the need to apply attentional resources at a critical moment of the trial. That is, the tasks we used were designed to place demands on the respondent to intensely apply attention at the right time and in such a way that respondents would not perform better or worse based solely on construct-irrelevant factors such as their processing speed ability and/or ability to optimize speeded vs. accurate responding. To account for both processing speed and individual differences in speed–accuracy tendencies, some of these new and modified tasks only required *accurate* responding without requiring *quick* responding (i.e., reaction time was irrelevant). We also accounted for speed–accuracy interactions by including tasks that were adaptive such that the task got more difficult as participants performed better or easier as they performed worse, and for these tasks the dependent variable was how quickly they could respond while also maintaining a certain level of accuracy that was the same for all participants. We called this method the *toolbox approach* because the goal was simply to develop more, and better, tasks to add to the metaphorical toolbox of tasks used to measure individual differences in attention control.

In the first study using the toolbox approach (Draheim et al., 2021), we administered a combination of ten existing, modified, and novel tasks designed to measure attention control and a battery of other cognitive tasks to 401 participants from the Atlanta, GA community. We then performed several validation analyses and assessed ten attention tasks based on internal consistency, test–retest reliability, average intercorrelation to each other attention task, average loading onto a factor comprised of three attention tasks, and relationship to working memory capacity and fluid intelligence. We found that the accuracy-based and adaptive tasks were *much* better on average across these validation criteria than traditional attention tasks such as reaction time and difference score-based Stroop and flanker. The top-performing attention tasks had around five times as much reliable and predictive variance as the traditional Stroop and flanker tasks, and the adaptive versions of the Stroop and flanker tasks had around three times as much predictive variance. The adaptive and accuracy-based measures were also superior to the psychomotor vigilance task – a traditional reaction time-based task of sustained attention.

Another key finding from Draheim et al. (2021) was that attention control, either partially or even fully (depending on which tasks were used), accounted for the relationship between working memory capacity and fluid intelligence at the latent level, supporting our theoretical position that the ability to control one's attention is the causal basis of

the typically strong relationship between working memory capacity and various other higher-order cognitive behaviors (see Burgoyne & Engle, 2020; Burgoyne et al., 2022; Draheim et al., 2022; Shipstead et al., 2016; Tsukahara et al., 2020; also see Conway et al., 2021 and Rueda, 2018).[1]

The present study

The present study was both a replication and an extension of the original toolbox approach (Draheim et al., 2021). Given the implications of the findings, it was important to test whether the key results could be replicated in a new data set and with some changes that accounted for limitations in the original study. Further, we also endeavored to continue iterating and improving the attention control tasks included in the study, as being able to assess this construct reliably and validly is especially important given the state of research in this area.

The data analyzed in the present study were part of a larger data collection effort that took place between November 2020 and April 2022 at the Georgia Institute of Technology. The study at large included roughly 50 cognitive tasks administered over five sessions plus in-lab and online follow-ups not relevant to the present study. A summary of the study at large and a reference list of all prospective publications from this project can be found at https://osf.io/qbwem. The constructs of interest for the present study are working memory capacity, fluid intelligence, attention control, and processing speed.

The rationale for task inclusion and modifications to the attention control tasks from Draheim et al., (2021) are described at length in Appendix A along with the justification for including selective visual arrays as an indicator of attention control. To summarize, the present study included modified versions of several attention-control measures from Draheim et al., a novel location-selection version of the visual arrays task, a broader assessment of working memory capacity, and slight changes to the assessment of processing speed. Further, unlike in Draheim et al., processing speed

tasks were administered to the full sample instead of a subset of the original sample.

Method

General procedure & participants

Participants completed a battery of tasks divided into four sessions, each designed to take approximately 2 h, but participants were informed that their participation could take as long as 2.5 h and to plan accordingly. Participants were recruited from several colleges in the Atlanta, GA area and the broader Atlanta community. To be eligible for the study, participants had to report being a fluent and native (learned at age 5 or earlier) English speaker aged 18–35 years with normal or corrected-to-normal vision, no history of seizure, and could not have participated in a previous study in our lab. A total of 327 participants completed the first four sessions. The final sample was $N = 301$, as described below in the data cleaning and processing procedures.

Participants were given a questionnaire asking about various demographic and background information and were informed that their answers were both anonymous and voluntary. Of the 301 participants included for final data analysis, 295 reported their age, gender, race, current college status, and languages spoken. The mean age was 21.95 years old (SD = 4.06). For college status, 70.2% reported being an undergraduate student, 17.6% graduate student, and 12.2% having no college-level education. Specifically, 45.4% were current students at Georgia Tech, 23.1% at Georgia State University, 6.1% at Kennesaw State University, 4.4% at Emory University, 8.8% at another institution. For gender, 58% reported identifying as female, 40.3% male, 1.4% self-identify/other, and 0.3% transgender. For race, 42% reported identifying as Asian or Pacific Islander, 29.5% White/Caucasian, 16.3% Other/Mixed-race, and 12.2% Black/African American. Finally, 68.5% reported English as their first language and 59.7% reported speaking another language (101 = fluently, 69 = somewhat fluently, and 6 = not very fluently).

Participants received $200 compensation for their participation in the study at large, distributed as $30 for the first session and each subsequent session providing $5 more than the previous. Georgia Tech students were given the option for each session to receive 2 h of participation credit for their psychology course(s) in lieu of monetary compensation. Participants could schedule sessions within designated slots during lab hours with the restriction that they could not complete more than one session per day. Due to COVID-19 restrictions, participants were run in separate rooms from each other, and our lab space was altered to accommodate up to four simultaneous participants for sessions 1–4 and an additional three for session

---

[1] Draheim et al. (2021) involved the administration of ten existing, modified, and newly developed tasks believed to measure attention control, including traditional reaction time-based versions of the Stroop, flanker, and psychomotor vigilance tasks. We then created 120 different three-task attention control factors using all possible combinations of tasks. Some, but not most, of these 120 attention control factors resulted in a full mediation of the working memory capacity–fluid intelligence relationship, and full mediation was only found with attention control factors that included the visual arrays task. Some attention control factors without the visual arrays task also explained all but 4–6% of the shared variance between working memory capacity and fluid intelligence at the latent level.

5. An experimenter was stationed at a desk just outside of the rooms, observing performance to ensure participants were following instructions. The experimenter also took notes regarding participant behavior, alertness, and apparent motivation, answered task-related questions, and started the run files for each task. Participants were not explicitly informed that they would be observed but they were aware of the experimenter's presence. Undergraduate and post-baccalaureate research assistants usually served as the experimenter, with graduate students and post-docs filling in as needed. At least one senior lab member (graduate student or post-doc) was present to supervise data collection. Longer tasks had built-in rest periods appear after a set number of trials designed to roughly occur approximately every 10–15 min. Participants could advance the rest screen at their convenience to continue performing the task. Participants were asked to avoid getting out of their chair while in the middle of a task if possible but were encouraged to take short breaks between tasks as needed.

The Georgia Institute of Technology Institutional Review Board approved the protocols (H20165 & H20532) and amendments for this study throughout data collection. We also followed Georgia Tech's COVID-19 guidelines and protocols, including COVID-19 screening for participants, temperature checks, mask-wearing, running participants in separate rooms along with other physical distancing procedures, and sanitizing rooms and equipment after each participant finished their session. Participants were not allowed into the building if they answered "yes" to any question on a COVID-19 checklist for possible symptoms and risk factors. Lab personnel were also strongly encouraged to get tested weekly and vaccinated. Lab personnel and recent participants were asked to report if they had a recent COVID-19 exposure or positive test. Data collection was temporarily halted on several occasions after a self-reported positive COVID-19 test or exposure from a participant or lab member.

Tasks of interest

All tasks included in the present study were programmed and run using E-Prime software. Task order was identical for every participant. Attention control and working memory capacity tasks were interspersed throughout sessions 1–4 such that each of the first four sessions had at least one task of each construct. The fluid intelligence tasks were in session 1–3. All three processing speed tasks were administered in the fourth session.

Due to adhering to COVID-19 protocols and various participant-specific factors (e.g., showing up late, taking long breaks, having to redo practice and instructions multiple times), sessions often went longer than the < 2.5 h cited to the participants, resulting in some participants needing to leave a session before completing all tasks. When possible, participants completed any remaining tasks at the beginning of their following session, but some participants were still unable to complete all tasks in the allotted time. When this happened, the tasks most likely to have missing scores were the processing speed tasks given they were the final tasks administered in the fourth session.

Attention control

Most attention tasks had practice-to-criterion requirements such that participants needed to perform at a certain level in the (easier) practice trials to advance to the critical trials. If they did not, they had to go through the instructions and practice again. If participants failed to reach the practice performance criterion of a task multiple times, a yellow screen would appear to alert the experimenter. The experimenter would attempt to explain the task to the participant again and see if they understood it, but in some cases the task was skipped due to either time considerations or because the participant appeared unmotivated or otherwise could not understand the instructions as explained by the experimenter.

Antisaccade (Hallett, 1978; Hutchison, 2007; Kane et al., 2001) Participants saw a central fixation cross lasting for 1000 or 2000 ms followed by an alerting tone for 300 ms. After the alerting tone, an asterisk appeared for 100 ms at 12.3° visual angle to the left or the right of the central fixation followed immediately by a target "Q" or an "O" for 100 ms on the *opposite* side of the screen from the asterisk. The location of the asterisk and target letter were both visually masked for 500 ms by "##". The participants' goal was to ignore the distracting asterisk and instead look away to the other side of the screen to catch the target "Q" or "O". Participants had as much time as needed to respond to which letter appeared by pressing the associated key on the keyboard. After responding, accuracy feedback was displayed for 500 ms, followed by a blank inter-trial interval of 1000 ms. Participants completed 16 practice trials with a slower presentation rate and then 72 test trials. The dependent variable was the proportion of correctly identified target letters.

Sustained attention-to-cue task (SACT; Draheim et al., 2021) This was an updated version of a novel task from Draheim et al. (2021) that was designed to be an accuracy analog of the psychomotor vigilance task. The main differences from the Draheim et al. (2021) version were, (1) the visual cue no longer remained on the display during the wait time interval, (2) there was no distractor presented just prior to stimulus onset, and (3) the array of target and non-target letters were made to be less uniform across trials.
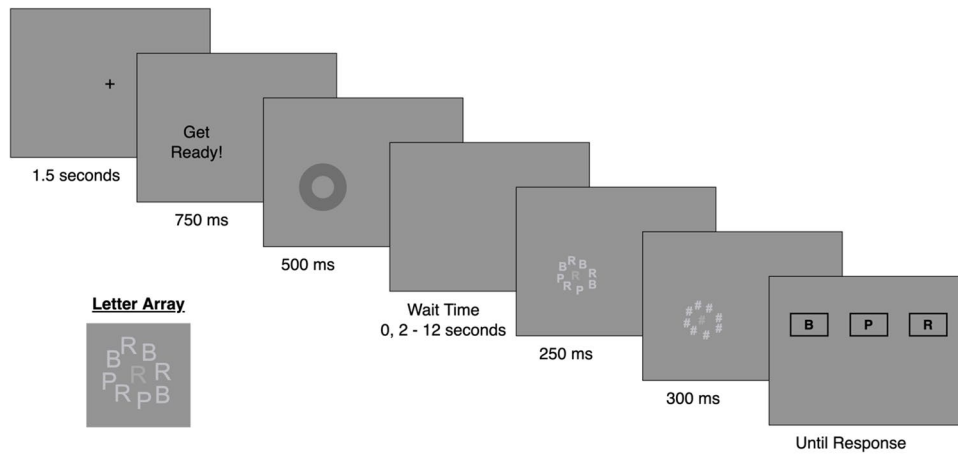
Fig. 1 Sustained attention-to-cue (SACT) task

The critical element in this task is the wait time interval in which attention must be sustained at a spatially cued location for a variable amount of time (Fig. 1). After the variable wait time, a target letter was briefly presented that needed to be identified amidst a mix of other non-target letters. Each trial started with a central black fixation for 1 s followed by a 750-ms interval in which the words "Get Ready!" were displayed at the to-be cued location along with an auditory beep. A circle cue was then displayed for approximately 500 ms, and then was removed from the display during the wait time interval. The wait time lasted either 0 s or 2–12 s in 500-ms intervals (e.g., 2, 2.5, 3, 3.5… s). After the variable wait time, an array of letters was displayed at the cued location for 250 ms and then masked for 300 ms. The target letter was identifiable as the central letter in slightly darker font color. The target and non-target stimuli were B, P, or R's. Responses were made using the mouse and selecting from boxes with a B, P, or R. The task had three blocks of 22 trials for a total of 66 trials without feedback. The dependent variable was the proportion of correct responses.

Color selective visual arrays (Color VA; Luck & Vogel, 1997; Shipstead et al., 2014) Participants saw an array with an equal number of blue and red rectangles in various orientations (horizontal, vertical, 45° diagonal or 135° diagonal; Fig. 2). Participants viewed a fixation cross for 800 ms and then the participant was cued to attend to either the red or blue rectangles by a 250-ms flash of "RED" or "BLUE". After a 100-ms delay, the array was presented for 250 ms and, after another delay of 900 ms, the array was presented again with only the rectangles of the target color. One of these rectangles was probed with a white dot and had changed orientation on 50% of the trials. The participant was asked whether the probed rectangle was in the same orientation as it was in the initial array, and responses were made by pressing the "S" (same) or "D" (different) key on the keyboard. Participants could take as long as needed to respond, and the next trial began 500 ms after each response.

Half the trials were of set-size 3 (three targets and three distractors) and half were of set-size 5. The different set sizes were interspersed within the same block of trials and
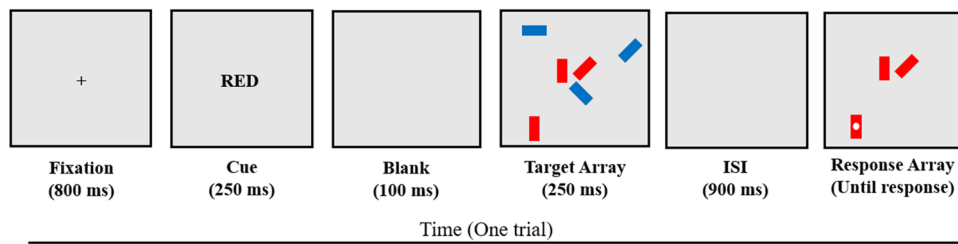

Fig. 2 Color selective visual arrays (Color VA). *Note.* Participants see six or ten total rectangles, half blue and half red. The participant is cued to only attend to rectangles of a particular color. After the target array and a 900-ms ISI, the rectangles of the cued color are re-presented, one of which has a white dot on it. The participant's goal is to indicate whether this probed rectangle is the same orientation as in the initial array, and the participant is asked to respond "Yes" if the rectangle is the same orientation and "No" if the rectangle has changed. In the trial shown, the participant should indicate "Yes". Example shown is set size 3. Figure is not to scale
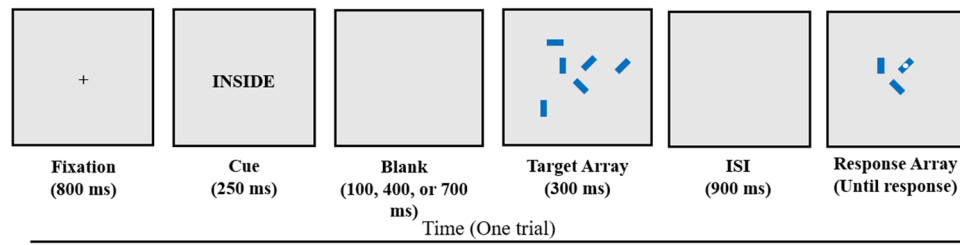
Fig. 3 Location-selective visual arrays (concentric VA). *Note.* Participants see six or ten blue rectangles, half of which occur at the center of the screen and half on an imaginary ring 9.3° from the center. The participant is cued to only attend to rectangles either at the center or on the periphery by cues of "INSIDE" or "OUTSIDE". After the target array and a 900 ms-ISI, the rectangles in the cued location are re-presented, one of which has a white dot on it. The participant's goal is to indicate whether this probed rectangle is the same orientation as in the initial array, and the participant is asked to respond "Yes" if the rectangle is the same orientation and "No" if the rectangle has changed. In the trial shown, the participant should indicate "Yes". Example shown is set size 3. Figure is not to scale

participants were not aware which set size would occur on any given trial. The task consisted of 120 total test trials, 60 of set-size 3 and 60 of set-size 5.

The dependent variable was a capacity score (k) calculated using the single probe correction (see Cowan et al., 2005; Shipstead et al., 2014). This calculation is $N *$ (hits + correction rejections – 1), where N is the set size for that array. This calculation is done separately for each set size and then the k scores from each set size are averaged.

Location-selective visual arrays (concentric VA) This was a novel version of the color VA task described above (see Appendix A). Participants saw an array of 6 or 10 small blue rectangles in various orientations (horizontal, vertical, 45° diagonal or 135° diagonal; Fig. 3), with half semi-randomly assorted inside a small (3° visual angle) imaginary circle at the center of the screen and the other half appearing along an imaginary ring of 9.3° visual angle, each randomly presented in one of 12 equally spaced locations that corresponded to hours on a clock. Participants viewed a fixation cross for 800 ms and then were cued to attend to rectangles that were inside the focal circle or on the outside ring via a 250-ms flash of either "INSIDE" or "OUTSIDE." After a 100-, 400-, or 700-ms delay, the array was presented for 300 ms and, after another delay of 900 ms, the array was presented again with only the rectangles in the cued region. One of these rectangles was probed with a white dot and had changed orientation on 50% of the trials. The participant was asked whether the probed rectangle was in the same orientation as it was in the initial array, and responses were made by pressing the "S" (same) or "D" (different) key on the keyboard. Participants could take as long as needed to respond, and the next trial began 500 ms after each response.

Half the 120 test trials were of set-size 3 (three targets and three distractors) and half were of set-size 5. The

different set sizes were interspersed within the same block of trials and participants were not aware which set size would occur on any given trial. The dependent variable was calculated as a k score as described above for the color VA task.

Several measures were taken to help ensure participants understood the task and could differentiate between stimuli that were focal (within the imaginary circle) and peripheral (on the imaginary ring outside the center of the screen). This included extensive piloting, modifications during piloting, and more involved instructions and practice than the other tasks in the study.

Color Stroop with adaptive response deadline (Stroop DL; Draheim et al., 2021; Stroop, 1935) This task was a modified version of the one used in Draheim et al. (2021). In this version of the task, the adaptive procedure occurred on a trial-by-trial basis but only for incongruent trials. In the version used in Draheim et al., the adaptive procedure occurred after a block of 16 trials and was based on performance from both congruent and incongruent trials.

The words "RED", "GREEN", and "BLUE" were presented one at a time in red, green, or blue font colors. The words were either congruent with the color (e.g., the word "RED" in red font color) or incongruent with the color (e.g., the word "RED" in blue font color). There was a 2:1 ratio of congruent to incongruent trials with 96 incongruent trials and a total of 288 trials overall. The task was administered over four blocks of 72 trials.

An adaptive staircase procedure was used to estimate the subject's response deadline threshold, based solely on incongruent trial performance. On each incongruent trial, if an incorrect response was made or the reaction time was longer than the response deadline then the response deadline increased (more time to respond) on the next trial. If a correct response was made and the reaction time was shorter than the response deadline, then the

response deadline decreased (less time to respond) on the next trial. The response deadline started at a relatively easy level of 1500 ms. A 3:1 up-to-down ratio was used for the step sizes such that the step size (change in response deadline) for incorrect/too slow of trials was three times larger than the step size for correct/deadline met trials. The step size started at 240:80 ms, decreased to 120:40 ms after 17 incongruent trials, decreased to 60:20 ms after 33 incongruent trials, decreased to 30:10 ms after 49 incongruent trials, decreased to 15:5 ms after 65 incongruent trials, and finally settled at 9:3 ms after 81 incongruent trials. Feedback was given in the form of an audio tone and the words "TOO SLOW! GO FASTER!" presented in red font when the response deadline was not met. The dependent variable was the average response time of the final four reversals, a reversal being a trial in which the deadline either increased after it had decreased on the previous trial or decreased after it had increased.

Arrow flanker with adaptive response deadline (flanker DL; Draheim et al., 2021; Eriksen & Eriksen, 1974; Nieuwenhuis et al., 2006; Stoffels & van der Molen, 1988)

This task was a modified version of the one used in Draheim et al. (2021). In this version, the adaptive procedure occurred on a trial-by-trial basis but only for incongruent trials. In the version used in Draheim et al., the adaptive procedure occurred after a block of 16 trials and was based on performance from both congruent and incongruent trials.

A target arrow was presented in the center of the screen pointing either left or right along with two flanking arrows on either side. The flanking arrows were either all pointing in the same direction as the central target (congruent trial) or all in the opposite direction (incongruent trial). There was a 2:1 ratio of congruent to incongruent trials with 96 incongruent trials and a total of 288 trials overall. The task was administered over four blocks of 72 trials each with an optional rest break between blocks.

An adaptive staircase procedure was used to estimate the subject's response deadline threshold. The adaptive procedure was based only on the incongruent trials. On each incongruent trial, if an incorrect response was made or the reaction time was longer than the response deadline, then the response deadline increased (more time to respond) on the next trial. If a correct response was made and the reaction time was shorter than the response deadline, then the response deadline decreased (less time to respond) on the next trial. The response deadline started at a relatively easy level of 1500 ms. A 3:1 up-to-down ratio was used for the step sizes such that the step size (change in response deadline) for incorrect/too slow of trials was three times larger than the step size for correct/deadline met trials. The step size started at 240:80 ms, decreased to 120:40 ms after 17

incongruent trials, decreased to 60:20 ms after 33 incongruent trials, decreased to 30:10 ms after 49 incongruent trials, decreased to 15:5 ms after 65 incongruent trials, and finally settled at 9:3 ms after 81 incongruent trials. Feedback was given in the form of an audio tone and the words "TOO SLOW! GO FASTER!" presented in red font when the response deadline was not met. The dependent variable was the average response time of the final four reversals, a reversal being a trial in which the deadline either increased after it had decreased on the previous trial or decreased after it had increased.

Working memory capacity

Working memory capacity was assessed using two different paradigms – running span (numbers and digits) and complex span (symmetry and rotation). The complex-span tasks consist of alternating memory storage and processing sub-tasks (Conway et al., 2005; Unsworth et al., 2005). The advanced versions of the tasks include larger set sizes of memory items (Draheim et al., 2018) and two blocks of each set size rather than three. The running-span tasks consisted of maintaining and updating a specified number of items in a continuous stream of presented memory items. For the dependent variable in all working memory capacity tasks, we used the edit-distance scoring method to calculate span scores (Gonthier, 2022). This method is a recently proposed alternative method shown to be equally reliable but more precise than the traditional partial score, which requires items to be correctly recalled in their absolute serial position (Conway et al., 2005).

Advanced symmetry span (Draheim et al., 2018; Unsworth et al. 2009)

This task required participants to remember a series of spatial locations in a 4x4 matrix presented in alternation with a pattern of squares which they had to decide whether the pattern was symmetrical on the vertical midline. On each trial, participants were first presented with a 16x16 matrix of black and white squares and were required to decide whether the pattern was symmetric on the vertical midline. Followed by the symmetry judgment, a 4x4 matrix of squares with one square highlighted in red were displayed. The location of the red square was the to-be-remembered spatial location. This alternation continued until a variable set-size of spatial locations had been presented. Then, on the recall screen the same 4x4 matrix of squares was presented but with no squares highlighted in red. Participants had to recall the spatial locations in the correct order by clicking the mouse on the appropriate squares in the matrix. There were a total of 12 trials (two blocks of six trials), set sizes ranged from 2–7, and each set size occurred twice (once in each block).

Advanced rotation span (Draheim et al., 2018; Kane et al., 2004) This task required participants to remember a series of directional arrows of varying size in alternation with a mental rotation task in which they had to mentally rotate and decide if a letter was mirror reversed or not. On each trial, participants first solved a mental rotation problem followed by the presentation of a single arrow with a specific direction (eight possible directions; the four cardinal and four ordinal directions) and specific size (small or large). Both the direction and size of the arrow were the to-be-remembered features. This alternation continued until a variable set size of arrows had been presented. Then, on the recall screen all possible arrow directions and sizes were presented. Participants had to recall the direction and size of the arrows in the correct order by clicking the mouse on the appropriate arrow. There were a total of 12 trials (two blocks of six trials), set sizes ranged from 2–7, and each set size occurred twice (once in each block).

Running letter span (Broadway & Engle, 2010; Pollack, et al., 1959) A series of 5–9 letters were presented each for 300 ms followed by a 200-ms delay. Participants were asked to recall the last 3–7 letters, depending on set size. Participants were informed of how many items they would need to remember at the beginning of a block of three trials, with blocks randomly presented. There were a total of 15 trials.

Running digit span (Cowan et al., 2005; Pollack, et al., 1959) A series of 12–20 digits were presented through headphones at a rate of four digits per second. Participants were asked to recall the last six digits, and 18 trials were presented.

Fluid intelligence

Raven's advanced progressive matrices – odd problems (Kane et al., 2004; Raven & Raven, 2003) Participants were presented with a matrix of figures that follow a logical pattern across rows and columns. For each problem in this task, a 3x3 matrix of eight abstract figures was presented with the bottom-right element missing. Participants had to identify the logical pattern and select one of eight answer choices that fits the logical pattern of the matrix. Participants were given 10 min to solve 18 of the odd-numbered problems from the full test. The dependent variable was the total number of problems solved correctly.

Letter sets (Ekstrom et al., 1976) Participants were presented with five sets of four-letter sequences (e.g., "NOPQ DEFL ABCD HIJK UVWX") and were asked to identify a common pattern among four of the sets and select the set of letters that did not follow the pattern (e.g., the letter sets are all in consecutive alphabetical order except for DEFL). Participants were given 10 min to solve 30 problems. The dependent variable was the total number of problems solved correctly.

Number series (Unsworth et al., 2009; Thurstone, 1938) For each problem in this task, a series of numbers were presented that progressed in a particular logical fashion. Participants had to identify the rule and select the next number, out of five answer choices, which should occur next in the series of numbers to be consistent with the logical rule. Participants were given 5 min to complete 15 problems. The dependent variable was the total number of problems solved correctly.

Processing speed

All processing speed measures were computerized versions of paper-and-pencil tests. Participants were instructed to respond as quickly and accurately as possible, but consistent with standard administration procedures, were not alerted of the time limits of each task in the instruction phase. In each processing speed task, a pair of stimuli presented on the left and right side of the center of the screen needed to be compared and judged as to whether they are identical or not. All processing speed tasks were scored as the number of total points, with inaccurate responses counted as – 1 point and accurate responses at + 1 point.

Letter string comparison (Redick et al., 2012; Salthouse & Babcock, 1991) In this version of the letter string comparison task, participants viewed strings of three, six, or nine consonants appearing to the left and right of the center of the screen. The letter strings could either be the same or differ by a single letter. If different, the mismatching letter could appear in any location in the string. Participants indicated their response by clicking on a button on the screen labeled SAME for identical strings or DIFF for mismatching strings. Letters were printed in white size 18-pt. Courier New font on a black background. After completing six practice trials, participants completed two 30-s blocks of the task.

Digit string comparison (Redick et al., 2012) In this version of the digit string comparison task, participants viewed strings of three, six, or nine digits appearing to the left and right of the center of the screen. The digit strings could either be the same or differ by a single digit. If different, the mismatching digit could appear in any location in the string. Participants indicated their response by clicking on a button on the screen labeled SAME for identical strings or DIFF for mismatching

strings. Digits were printed in white size 18-pt. Courier New font on a black background. After completing six practice trials, participants completed two 30-s blocks of the task.

Pattern comparison (Redick et al., 2012; Salthouse & Babcock, 1991) In this version of the pattern string comparison task, participants viewed two patterns of line segments appearing to the left and right of the center of the screen. The patterns were three, six, or nine line segments connected to form simple to complex patterns. The patterns could either be the same or differ by a single line segment. If different, the mismatching line segment could appear in any location on the pattern. Participants indicated their response by clicking on a button on the screen labeled SAME for identical patterns or DIFF for mismatching patterns. Patterns were printed in black against a white background. After completing six practice trials, participants completed two 30-s blocks of the task.

Data processing A total of 327 participants completed the first four sessions. Data processing, cleaning, scoring, and analyses were conducted with R statistical software (R Core Team, 2020) and EQS software. The visual arrays and complex-span task scores were calculating using the *englelab* (Tsukahara, 2022) R package, exploratory factor analysis was conducted using the *psych* (Revelle, 2022) R package, and structural equation models were conducted using the *lavaan* (Rosseel, 2012) R package. For all tasks, except fluid intelligence, scores were labeled as missing if they were at or below chance performance (for complex span tasks this was based on the processing, not storage, part of the task; e.g., symmetry judgments for symmetry span). For all tasks, a two-pass outlier method was used on scores. On each pass, z-scores were computed and univariate outliers were identified as having scores $\pm$ 3.5 standard deviations or greater from the mean score on that pass. Outlier scores were also relabeled as missing. Data from participants who had missing scores on more than three out of six attention control tasks, more than one fluid intelligence task, any working memory capacity tasks, and/or more than one processing speed task were removed from the final dataset, which resulted in the removal of data from 26 participants and a final sample of $N = 301$. This final dataset therefore consisted of 301 participants who met the criteria for data inclusion, and any missing scores from these participants (2.1% of total task scores) were then imputed with expectation maximization using EQS software. Monte Carlo simulations suggest that sample sizes should approach around 250 to obtain stable estimates of correlations (Schönbrodt & Perugini, 2013).

Results

Task level

Descriptive statistics, reliabilities, and proportion of missing (and, therefore, imputed) scores for each task are shown on Table 1. The tasks were generally normally distributed except for the adaptive Stroop and flanker tasks, which had a positive skew and large kurtosis. Internal consistency calculations ranged from .73 to .94, with all attention control tasks at or above .79.

With one exception, the attention control tasks all had statistically significant intercorrelations at the task level as well as significant correlations with factor scores for attention control, working memory capacity, fluid intelligence, and processing speed (Table 2; see Table 5 for the full correlation matrix). The exception was flanker DL, which had weak/null correlations to the other attention control tasks, a weak correlation with the attention control factor score, and non-significant correlations with factor scores for the other cognitive constructs included in the study.

The intercorrelations among the attention control tasks were moderate in strength with an average of $r = .30$, or 9% shared variance at the task level (Table 3). This was affected by the inclusion of two similar tasks (visual arrays) but also by the inclusion of the flanker DL task, which had very weak correlations with all other tasks. Removing the flanker DL from analysis resulted in slightly stronger intercorrelations among the tasks, with all measures correlated above $r = .30$ and an average intercorrelation of $r = .39$ (15% shared variance).

Factor structure

To test the factor structure of the included measures, we ran a planned oblique factor analysis with all tasks of interest. Based on advice from reviewers, we evaluated several metrics to determine the number of factors: eigenvalues greater than 1, parallel analysis with 1000 iterations (Horn, 1965), and Bayesian information criterion (BIC) model comparison. Five factors had an eigenvalue greater than 1, parallel analysis suggested a five-factor solution, and a five-factor solution had the smallest BIC (4-factor: BIC = – 176.82, 5-factor: BIC = – 189.82, 6-factor: BIC = – 151.64). Therefore, a five-factor solution was used. Factors comprised of, in order of eigenvalues, the attention control tasks, the processing speed tasks, the running span tasks, the complex span tasks, and the fluid intelligence tasks (Table 4). All tasks loaded at or above .30 onto their respective factors and below .30 onto all other factors.

This exploratory factor analysis informed a follow-up confirmatory factor analysis in which each task was loaded

Table 1 Descriptive statistics

Task	Mean (SD)	Skew	Kurtosis	I.C.	Missing (%)
Antisaccade	.81 (.12)	− 0.60	− 0.65	.87[a]	5.7
VA – color	2.5 (0.7)	− 0.41	− 0.21	.81[b]	0.7
VA – concentric	2.4 (0.7)	− 0.67	− 0.01	.79[b]	0.7
SACT	.89 (.10)	− 1.15	0.92	.87[a]	4.0
Stroop DL	1023 (484)	1.63	2.53	.87[a]	3.0
Flanker DL	659 (270)	1.81	2.88	.89[a]	3.3
Symmetry span	35.1 (8.8)	− 0.49	− 0.09	.74[a]	3.0
Rotation span	29.2 (8.5)	− 0.40	− 0.01	.76[a]	3.0
Running span - digits	66.9 (12.2)	− 0.33	1.11	.83[a]	2.3
Running span - letters	51.6 (10.1)	− 0.08	− 0.26	.75[a]	3.3
Raven's advanced	11.4 (2.8)	− 0.38	− 0.26	.77[a]	1.3
Letter sets	16.5 (4.4)	− 0.13	− 0.66	.85[a]	2.0
Number series	10.0 (3.0)	− 0.24	− 0.70	.73[a]	1.0
Digit comparison	29.9 (5.5)	− 0.49	0.08	.88[b]	0.0
Letter comparison	20.6 (4.1)	0.12	0.45	.82[b]	0.0
Pattern comparison	39.1 (6.0)	− 0.09	− 0.19	.94[b]	0.7

I.C. = Internal consistency, calculated before imputation. [a] = assessed using Cronbach's alpha; [b] = assessed using split-half and stepped up with the Spearman–Brown prophecy formula. Missing % = total number of missing cases from the sample of $N = 301$, which were then imputed. Antisaccade and SACT scores are the proportion of correct responses; VA tasks are capacity (k) scores; Stroop DL and flanker DL scores are the average response time (ms) of the last four reversals in the adaptive procedure; symmetry span, rotation span, and running span tasks are edit distance scores for number of complete responses in the correct position (see Gonthier, 2022); all other tasks are total number of correct trials

Table 2 Attention control task correlations

Task	VA-color	VA-conc.	SACT	Stroop DL	Flanker DL	AC	WMC- complex	WMC – running	Gf	PS
Antisaccade	.41	.42	.35	.35	.08 (*ns*)	.61	.39	.31	.27	.42
VA-color	-	.72	.35	.28	.12	.85	.60	.28	.48	.37
VA-conc.		-	.41	.36	.13	.92	.50	.30	.39	.32
SACT			-	.24	.09 (*ns*)	.56	.22	.25	− .04 (*ns*)	.30
Stroop DL				-	.11	.46	.25	.26	.18	.35
Flanker DL					-	.21	.03 (*ns*)	.00 (*ns*)	− .04 (*ns*)	− .05 (*ns*)

VA = visual arrays; SACT = sustained attention-to-cue; DL = adaptive response deadline; AC = attention control; WMC = working memory capacity; Gf = fluid intelligence; PS = processing speed. Intercorrelations among the attention control measures along with correlations between each attention control task and the cognitive constructs included in the present study. The construct correlations are regression factor scores obtained from the same exploratory factor analysis shown in Table 4. For ease of interpretation, some correlations were multiplied by − 1 such that positive correlations reflect that better performance on one task was associated with better performance on the other task or construct. All correlations above were statistically significant at α = .05 unless noted as "*ns*"

Table 3 Average intercorrelation among attention control tasks

Task	Average intercorrelation	With flanker DL removed
VA-concentric	.41	.48
VA-color	.38	.44
Antisaccade	.32	.38
SACT	.29	.34
Stroop DL	.27	.31
Flanker DL	.11	-
Total average	.30	.39

onto the factor it most strongly preferred in the exploratory factor analysis (Fig. 4). We had originally planned to load all four working memory capacity tasks onto one factor, however this model had very poor fit. Creating two separate working memory capacity factors then loading them onto a common working memory capacity factor also did not work due to issues with model fit and convergence. Given this and the exploratory factor analytic results, the working memory capacity tasks were loaded onto separate factors and no common factor was created.

There were also two changes for the attention control factor in the confirmatory model. The flanker DL task was

Table 4 Exploratory factor analysis – rotated pattern matrix

Task	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Antisaccade	**.46**	.18	.08	.04	– .03
Visual arrays - color	**.67**	.00	– .03	.14	.14
Visual arrays - concentric	**.82**	– .05	.02	.01	.07
SACT	**.54**	.17	.11	– .06	– .29
Stroop deadline	**.36**	.18	– .09	– .03	– .06
Flanker deadline	**.30**	– .09	– .01	– .12	– .05
Symmetry span	– .02	.02	.02	**.73**	.06
Rotation span	.03	.00	.05	**.74**	– .09
Running span – digits	.00	.07	**.79**	.02	.01
Running span – letters	.04	– .09	**.77**	.04	.04
Raven's advanced	.24	– .06	.07	.19	**.41**
Letter sets	– .04	.24	.26	– .03	**.52**
Number series	.13	.09	.04	.03	**.66**
Digit comparison	.09	**.75**	– .01	– .01	.06
Letter comparison	– .09	**.74**	.05	.05	.01
Pattern comparison	.17	**.43**	– .18	.24	.13

Extraction done via principal axis factoring; oblimin rotation. The strongest loading for each task is in bolded font. For ease of interpretation, some loadings were multiplied by –1 such that positive loadings reflect better performance on the task was positively related to the factor. Factor correlations were between $r = .27$ and .42

removed because of its weak intercorrelations to the other attention control tasks (Table 2), and it had the weakest loading onto the attention control factor in the exploratory factor analysis. The concentric VA task was also removed. Even though this task had strong correlations with the other measures and loaded strongly with the other attention tasks, including both the concentric VA and color VA tasks resulted in a somewhat imbalanced attention control factor with very strong loadings for those two tasks and weak loadings for the other tasks (see Fig. 8). The color VA task was retained over the concentric VA task because it is a more typical visual arrays task and was very similar to the selective visual arrays task included in Draheim et al. (2021). The flanker DL and concentric VA tasks were also dropped entirely from subsequent analyses, resulting in an attention control battery consisting of antisaccade, color VA, Stroop DL, and SACT. In the confirmatory model shown in Fig. 4, all tasks had a loading of at least .45 onto its respective factors, with generally quite strong loadings across the board. Factor paths were also quite strong and ranged from .34 to .67.

Attention control's mediation of the working memory capacity – Fluid intelligence relationship

A key finding from Draheim et al. (2021) was that attention control fully mediated the relationship between
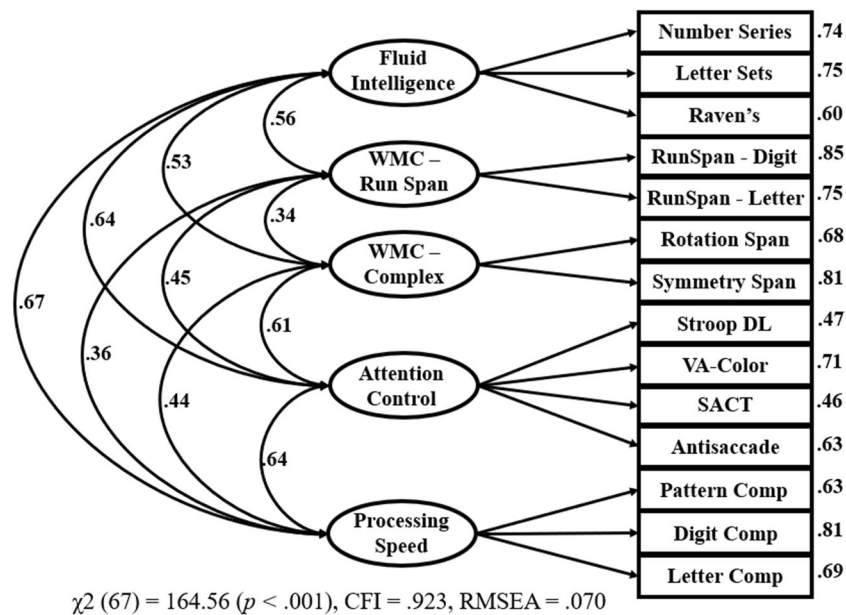

$\chi 2\ (67) = 164.56\ (p < .001)$, CFI = .923, RMSEA = .070

Fig. 4 Confirmatory factor analysis with all included tasks. *Note.* VA = visual arrays; SACT = sustained attention-to-cue; DL = adaptive response deadline; WMC = working memory capacity; Comp = Comparison. Factor loadings are shown on the right side of each task. For ease of interpretation, some loadings and paths were multiplied by – 1 such that positive values reflect that better performance on a task was associated with better performance on the other task or construct. All loadings and paths were statistically significant at $p < 0.05$
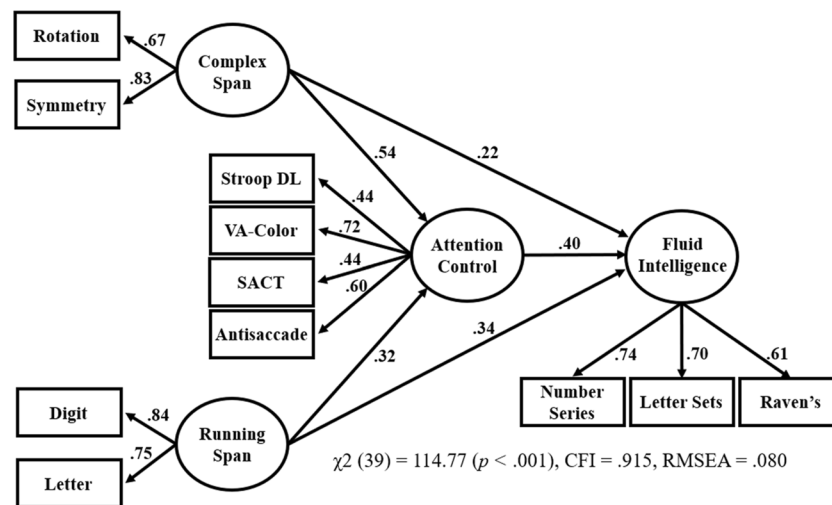
Fig. 5 Attention control mediating the working memory capacity–fluid intelligence relationship. *Note*. Structural equation model testing the magnitude of attention control's mediation of the working memory capacity–fluid intelligence relationship. VA = visual arrays; SACT = sustained attention-to-cue; DL = adaptive response deadline. Stroop DL's factor loading was multiplied by – 1 such that a positive value reflects that better performance on it loaded positively with the factor. All loadings and paths were statistically significant at the .05 level

working memory capacity and fluid intelligence when the strongest attention control measures (including visual arrays) from that study were used in the mediation model. We tested whether this finding was replicated in this present study by running a similar model, with two key differences in our approach in the present study. First, in Draheim et al., we tested all possible combinations of three-task factors and found that models using a combination of antisaccade, SACT, selective visual arrays, and the adaptive Stroop and flanker deadline tasks accounted for all, or nearly all, of the variance in the working memory capacity–fluid intelligence relationship, whereas in the present study we tested just one mediation model and with four attention control tasks. Second, Draheim et al. only included complex span (operation span, symmetry span, and rotation span) to measure working memory capacity, whereas the present study included running span. The factor analytic results below surprisingly revealed that running span and complex span tasks did not load together onto a common factor, which necessitated having them as separate factors in the mediation model tested here.

The critical mediation model is shown in Fig. 5. The path from attention control to fluid intelligence was stronger than the paths from the working memory capacity factors to fluid intelligence (.40 vs. .22/.34), but the paths from the two working memory capacity factors were still statistically significant after accounting for the role of attention control. To quantify the magnitude of the partial mediation, attention control accounted for 83% of the total variance in the complex span–fluid intelligence relationship (the path went from .53 to .22 after accounting for attention control) and 67% of the total variance in the running span–fluid intelligence relationship (the path went from .56 to .34). Note that in the present study, some three-task factors of attention control did result in a statistically complete mediation, for example when the attention control factor was comprised of antisaccade, color VA, and Stroop DL (see Fig. 9).

The role of processing speed

The final set of analyses tested whether the strong relationship observed between attention control and both working memory capacity and fluid intelligence could be attributable to construct-irrelevant factors, such as general speediness or motivation. It was recently shown that traditional attention control tasks are contaminated with performance variance associated with processing speed, response cautiousness, and/or motivation (e.g., Hedge et al., 2021), and this has been argued to be true for the attention control tasks used in our lab as well (e.g., Rey-Mermet et al., 2019; Rouder et al., 2022). To test this, the same structural equation model was used as in Draheim et al. (2021), which had both processing speed and attention control as correlated predictors of fluid intelligence and working memory capacity. If the attention control tasks are measuring processing speed, and not attention control, then we would not
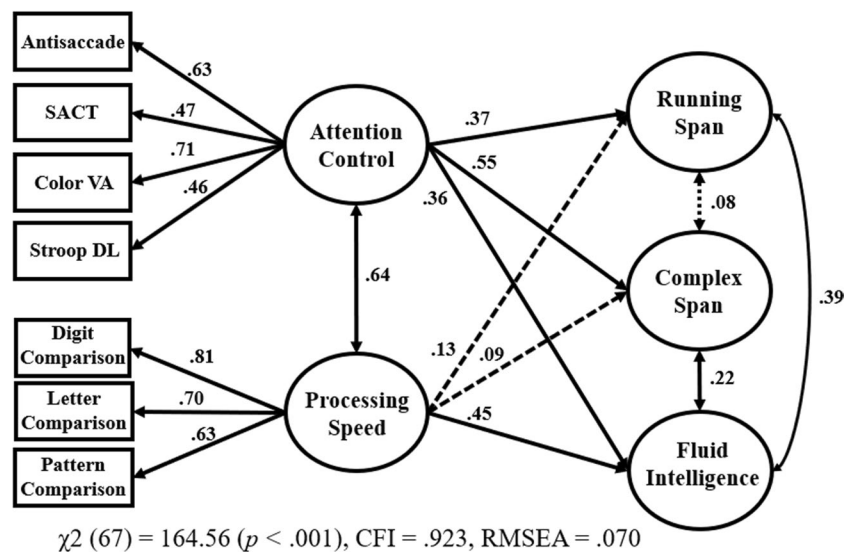
$\chi 2\ (67) = 164.56\ (p < .001)$, CFI $= .923$, RMSEA $= .070$

Fig. 6 Attention control and processing speed as correlated predictors of working memory capacity and fluid intelligence. *Note*. VA = visual arrays; SACT = sustained attention-to-cue; DL = adaptive response deadline; WMC = working memory capacity; Comp = Comparison. Factor loadings are shown on the right side of each task.

For ease of interpretation, some loadings and paths were multiplied by – 1 such that positive values reflect that better performance on a task was associated with better performance on the other task or construct. *Solid lines* indicate a statistically significant path ($p < 0.05$) whereas *dashed lines* indicate a statistically non-significant path

expect to see incremental validity of attention control to working memory capacity and fluid intelligence above and beyond the processing speed measures included in the study.

The model[2] is shown in Fig. 6. As in Draheim et al. (2021), the processing speed and attention control factors were strongly correlated (41% shared variance at the latent level), but attention control still had strong incremental prediction in working memory capacity and fluid intelligence above and beyond processing speed. Specifically, the path value from attention control to complex span was .55 (30% incremental variance explained above and beyond processing speed), the path value from attention control to running span was .37 (14% incremental variance), and the path value

from attention control to fluid intelligence was .36 (13% incremental variance). Conversely, processing speed's incremental prediction over attention control to complex span was .09 (< 1% incremental variance; *ns*), .13 for running span (2% incremental variance; *ns*), and .45 for fluid intelligence (20% incremental variance).

The pattern of results suggests that the commonality among the attention control and processing speed tasks was indeed the control of attention. The pattern of path values in the correlated predictors model (Fig. 6) were similar to a model in which we specified a common factor (that we attribute to attention control) and a unique processing speed factor (Fig. 7). Model comparisons showed that the common attention control-unique processing speed model (Fig. 7) was strongly preferred ($BF_{commonAC-uniquePS}$ = 33.76; $P(M_{commonAC-uniquePS} \mid Data) = .97$) to the correlated predictors model (Fig. 6). Also, the correlated predictor and common attention control-unique processing speed models were both strongly preferred ($BF_{correlatedpred}$ = 2717.20; $P(M_{correlatedpred} \mid Data) > .999$; $BF_{commonAC-uniquePS}$ = 91741.16; $P(M_{commonAC-uniquePS} \mid Data) > .999$) to a model with a common factor and a unique attention control factor (in which both the common and unique attention control factor were a significant predictor of the working memory capacity and fluid intelligence factors, see Supplemental Figure 3 for model details). In general, these models indicate that there is considerable common and unique variance in these attention control tasks with the processing speed tasks. These models do not entirely resolve the processing speed

---

[2] The attention control and processing speed latent factors are somewhat imbalanced in that there are four attention-control tasks and only three processing speed tasks. Also, the processing speed tasks share features with the attention tasks, thus possibly reducing their predictive validity due to also containing considerable method-specific variance. To address these issues, we (1) ran all combinations of the same model with only three of the attention control tasks, and (2) ran a model with a more diverse set of four processing speed tasks (see Supplemental Materials) and the four attention control tasks. This resulted in a total of five alternative correlated predictor models. Although there are some slight differences in model results, all the models showed that attention control predicts unique variance in the working memory capacity and fluid intelligence factors, but the processing speed tasks, by and large, do not. See Supplemental Figures 1 and 2 for model details.
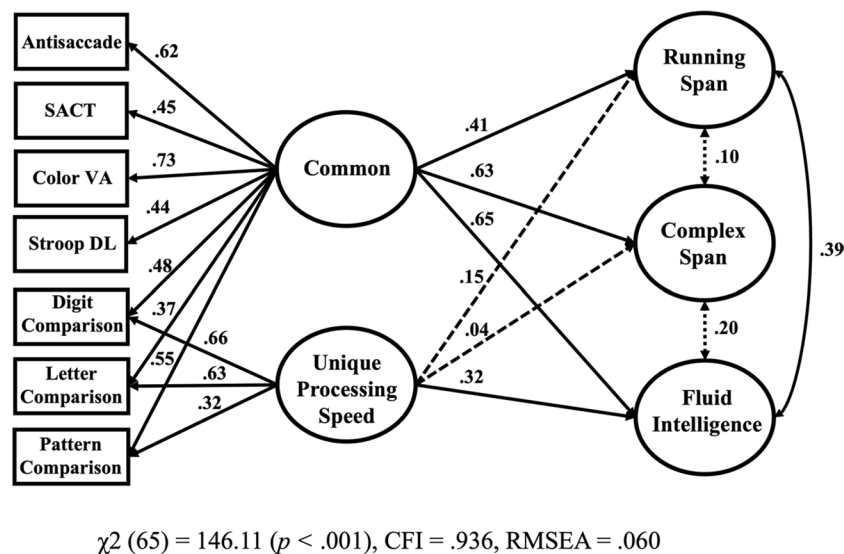
$\chi 2\ (65) = 146.11\ (p < .001)$, CFI $= .936$, RMSEA $= .060$

Fig. 7 The relationship between a common factor and a unique processing speed factor with working memory capacity and fluid intelligence. *Note*. VA = visual arrays; SACT = sustained attention-to-cue; DL = adaptive response deadline; WMC = working memory capacity. Factor loadings are shown on the right side of each task. For ease of interpretation, some loadings and paths were multiplied by – 1 such that positive values reflect that better performance on a task was associated with better performance on the other task or construct. *Solid lines* indicate a statistically significant path ($p < 0.05$)

vs. attention control debate (e.g., Rey-Mermet et al., 2019; Rouder et al., 2022), but they do challenge an anticipated interpretation of our data that the attention tasks used here only correlate with other cognitive abilities because they are assessing individual differences in something akin to processing speed.

Discussion

The present study was designed to extend and replicate the findings from Draheim et al. (2021). Specifically, we tested how strongly performance in two types of attention control tasks reflected a common demand for attention control, (1) accuracy-based measures that placed no demands on speeded responding, and (2) modified versions of Stroop and flanker tasks that had an adaptive response deadline. We also tested whether these attention control measures could fully account for the relationship between working memory capacity and fluid intelligence, and whether the predictive validity of these attention control tasks could be attributed to processing speed or general motivation. We found that (1) all but one of the attention control tasks included here were strong indicators of attention control in that they had good reliability, moderate intercorrelations, and strong loadings onto a single factor, (2) attention control was a very strong, though not statistically complete, mediator of the working memory capacity–fluid intelligence relationship, and (3) attention control predicted a substantial amount of

variance in working memory capacity and fluid intelligence above and beyond processing speed. Therefore, the present study replicated the findings from Draheim et al. It also extended the findings from Draheim et al. as the present study involved modifications to the attention control tasks designed to improve them from their first iterations, included a novel selective visual arrays measure, and had a broader assessment of working memory capacity.

Limitations

Although the present study largely replicated and extended the results of Draheim et al. (2021), there are limitations and caveats that warrant discussion.

The most salient difference in results from attention control tasks used in both studies was that, in Draheim et al. (2021) the flanker DL task was a strong measure of attention control according to our validation criteria, whereas in the present study this was not the case. The reason for this is not immediately apparent. The flanker DL task in the present study had a non-normal distribution of scores, but the same was also true of the Stroop DL. We performed some exploratory post hoc analyses to test whether the poor scores (i.e., the tail in the positive skew) in these tasks strongly influenced the results. Surprisingly, removing scores in the tail in the Stroop DL resulted in slightly stronger correlations to the other cognitive measures, but removing scores in the tail in flanker DL had no meaningful impact on its relationship to

the other tasks. That is, Stroop DL remained a decently strong measure of attention control regardless of whether data from the worst performers were included, whereas flanker DL performance still had little shared variance with the other tasks even when the worst performers were removed. Additional post hoc analyses also revealed that both the Stroop DL and flanker DL tasks were converging at a higher accuracy rate than intended, but also that, psychometrically speaking, the tasks were very similar aside from the difference in observed correlations involving them. We therefore recommend caution in using the deadline versions of these tasks, particularly flanker DL, pending more validation work on them.

Another limitation is that diversifying the measurement of working memory capacity by including verbal running span and spatial complex span tasks resulted in some complications. The running span and complex span tasks did not load onto same factor in latent analyses and, when loaded onto separate working memory capacity factors, these factors did not correlate strongly with one another. This is a surprising finding given that complex span and running span have been shown to be robust and strongly correlated measures of working memory capacity (Broadway & Engle, 2010; Conway et al., 2005). The lack of factor cohesion for these tasks necessitated running latent models with separate working memory capacity factors, which was not planned.

This also suggests the possibility that several of the latent factors in our models reflect considerable method variance rather than the construct of interest. For example, the running span tasks were verbal, the complex span tasks were spatial, the processing speed tasks are highly similar and only differ on the nature of the stimuli that are being compared (e.g., letters vs. digits), and the fluid intelligence tasks all require pattern recognition and completion. On the other hand, the attention control tasks do reflect a broader set of stimuli, response sets, and attentional demand (e.g., sustained attention in the SACT, response conflict in the Stroop DL, selective attention in the visual arrays).

Although the role of method variance may offer alternative explanations to our findings, there is a solid argument to the contrary. There is evidence to suggest that different types of working memory tasks do measure unique variance related to fluid intelligence (Shipstead et al., 2014; Shipstead et al., 2012, Redick and Lindsey, 2013). If they were simply capturing differences in method variance, then it would not be expected that they would show unique contributions to fluid intelligence. Therefore, even though we attempted to diversify our measure of working memory capacity, this may not have been entirely warranted in the first place. If this is the case, this suggests further research is needed to better understand the processes that underlie different working memory tasks. As for the processing speed tasks, we also conducted a correlated predictors model with attention control using a more diverse set of processing speed tasks

(see Supplemental Figure 2) and the general conclusions from the model did not change.

It is important to note that in correlational datasets with large sample sizes, effect sizes should be emphasized over the statistical significance of results, as most relationships will be statistically significant. In that lens, the magnitude of the mediated path from complex span to fluid intelligence (.22) was virtually identical to magnitude of mediation models run in Draheim et al., that did show a complete mediation. Also important is that in the present study, some mediation models specifying three indicators for the attention control latent did result in a full mediation of the complex span–fluid intelligence relationship in the present study (see Table 5). It was also the case in Draheim et al. that some mediation models resulted in full mediation and others partial mediation. Therefore, both studies demonstrated that when attention control is assessed with more reliable and valid tasks, attention control accounts for virtually all the variance in the relationship between complex span and fluid intelligence and the remaining, unexplained, variance in the relationship borders on statistical significance depending on which exact combinations of tasks are used to index attention control. However, the path from running span to fluid intelligence was larger, at .32, indicating that 10% of the variance in that relationship was unaccounted for by attention control. Why a relatively small proportion of the shared variance (around 4–10%) in the working memory capacity–fluid intelligence relationship is not *always* accounted for by these attention control tasks, depending on how the constructs are measured, is an open question for future research. One possibility is that of content validity, namely that the attention control tasks used here do not capture all relevant aspects of the broad ability of attention control. Another possibility is that both working memory capacity and fluid intelligence tasks require various strategies and mechanisms associated with retrieval from secondary memory, and that these attention control tasks do not place strong demands on secondary memory processes (e.g., Mogle et al., 2008; Unsworth et al., 2014; Unsworth & Spillers, 2010). There could also be other methodological explanations pertaining to properties of the working memory capacity and/or fluid intelligence tasks.

Attention control or processing speed?

One of the most difficult challenges in the behavior sciences is how to assess variables of interest reliably and validly. Unfortunately, it has often been shown that many cognitive measures are not as process pure as often assumed (see Draheim et al., 2019; Hedge et al., 2018, 2021; Miller & Ulrich, 2013; Wickelgren, 1977). The question of whether traditional attention control tasks measure individual differences in the ability to control attention has become contentious, and with little agreement on what constitutes evidence

of validity. Disagreements arise in part due to differences in emphasis placed on the use of contrast comparisons (i.e., difference scores), which is a method in experimental research to isolate *within*-subject effects. Some researchers view contrast comparisons as necessary and sufficient pieces of evidence for construct validity (e.g., process purity), even in individual differences research. We adopt a more comprehensive view of validity that requires converging evidence from multiple aspects of validity (Embretson, 1996). Experimental conditions that manipulate the demand on controlled processes can be an important aspect of validity but there has to be sufficient *between*-subject variability and smaller to more moderate correlations between conditions—two criteria that have rarely been met in past studies.

Regarding the assessment of attention control, the premise for researchers who favor using contrast comparisons is that baseline trials are necessary to account for construct-irrelevant sources of variance, and specifically that a failure to include such comparisons in attention control tasks will result in contamination from factors such as processing speed. Processing speed was discussed at some length in Draheim et al. (2021) and interested readers are encouraged to see that paper for the discussion. But two things are important to mention. The first is that contrast comparisons (i.e., difference scores) have consistently been shown to not be a viable method for eliminating variance attributable to processing speed. Ironically, using reaction time difference scores appears to artificially introduce additional variance associated with processing speed because respondents who are generally slower have larger difference scores independent of their cognitive ability (see Draheim et al., 2021; Hedge et al., 2021; Rey-Mermet et al., 2019; Verhaeghen & De Meersman, 1998; Zheng et al., 2000). For example, consider two respondents who are both twice as slow on incongruent Stroop trials than congruent ones. If one has an average reaction time on congruent trials of 500 ms and the other has an average reaction time of 1000 ms on congruent trials, their respective reaction time Stroop effects will also be 500 and 1000 ms. This would suggest that the former was much better at resolving cognitive conflict even though both participants had the same proportional change in reaction time when comparing their congruent and incongruent trial performance.

Second, we have shown that individual differences in performance on typical processing speed tasks do not account for the relationship attention control has with other cognitive ability measures (Draheim et al., 2021). Using other *between*-subjects measures that capture performance differences due to other constructs, such as processing speed, is an alternative to the use of difference score. The idea is that we cannot experimentally control for differences *between* individuals, but we can obtain measures to get at different sources of *between*-subject variation. If we can show that attention control predicts variance in abilities such as working memory capacity and fluid intelligence

after statistically controlling for performance differences on processing speed tasks, then this provides *a* piece of evidence for construct validity.

When we use this approach, we have found that processing speed and attention control tasks clearly load onto two separate factors in both exploratory and confirmatory factor analyses. This suggests that the two sets of tasks have sufficient unique variance between them that we can consider them as reflecting separate, yet related, mechanisms. Second, both this study and Draheim et al. (2021) showed that attention control had substantial incremental prediction of working memory capacity and fluid intelligence over processing speed. Third, the accuracy-based attention control tasks included in the present study placed no demands on speeded responding, as participants were afforded as much time as they needed to respond on each trial. Given this, it seems unlikely that tasks such as antisaccade, visual arrays, and/or SACT are simply processing speed measures masquerading as attention control tasks (cf. Rouder et al., 2022). It instead seems more plausible that processing speed tasks generally require a good degree of attention control, as individuals must sustain a high level of task engagement throughout the task while avoiding both internal and external distraction, for example micro- or macro-level lapses of attention in the form of mind wandering or attentional capture (e.g., Fukuda & Vogel, 2009). This may explain the strong latent correlation between the processing speed and attention control factors in the model shown in Fig. 6.

We are by no means arguing that this line of work represents the pinnacle of assessing individual differences in attention control. Rather, construct validity will necessarily be an ongoing development in this field as we gain advances in the measurement and theoretical understanding on the nature of attention control and, hopefully, some of the tasks outlined here are subject to further validation and modification from other researchers in different labs and populations. We would, however, caution the over-interpretation of null results from contrast comparisons in individual differences research whether analyzed using difference scores, bi-factor latent models, or hierarchical linear models—especially when there is little *between*-subject variability and strong correlations between conditions. Null results from these types of models do force us into an uncomfortable position of reconsidering how attention control should be assessed—and cognitive processes that give rise to individual differences in performance more generally. However, we strongly urge researchers to not abandon attention control as a valid psychological and psychometric construct and instead encourage the continued exploration into advances in the measurement and theoretical nature of attention control. In the present line of work, we have provided substantial evidence that attention control can be measured in a reliable and valid way. We also direct interested readers to a related study by Burgoyne

et al. (2023) who found evidence that attention control can be reliable and validly measured using modified versions of conflict tasks that can be administered in under 3 min. each.

Conclusion

A great deal of research has shown that measures of attention control which use difference scores and place demands on speeded responding, without accounting for differences in speed–accuracy tendencies, do not reliably assess individual differences in attention control. On the other hand, the present line of research demonstrates that accuracy-based measures in which reaction time is irrelevant can be both reliable and valid measures of this important ability. Most importantly, when attention control is measured more robustly, it accounts for virtually all the variance in the relationship between working memory capacity and fluid intelligence. This finding is congruent with research supporting the theoretical position that attention control, and not working memory capacity, is the primary driver of higher-order cognitive behavior (see Draheim et al., 2022). Using traditional and flawed tasks such as Stroop and flanker to assess individual differences in attention control therefore tend to bias results and conclusions by underestimating the relationship attention control has with other cognitive abilities. Our hope is that the field can soon move past assessing individual differences in attention control in this manner so that research efforts can be directed to more important theoretical questions and applications related to the importance of attention control as a marker for human cognition.

Appendix A

Rationale for task inclusion and changes from Draheim et al. (2021).

Attention control

The stronger half of the ten attention control tasks used in Draheim et al. (2021) were included in the present study: antisaccade, selective visual arrays, sustained attention-to-cue (SACT), adaptive arrow flanker with a response deadline (flanker DL), and adaptive color Stroop with a response deadline (Stroop DL). These tasks are described in detail in the Method section in the main body of this paper, and readers are also encouraged to see Draheim et al. for a full description of the previous version of each task. Standard versions of Stroop and flanker, or other traditional attention measures, were not included in this study as it has been repeatedly demonstrated that they have relatively poor reliability and validity in regard to assessing individual differences in attentional mechanisms (e.g., Friedman & Miyake, 2004; Paap & Sawi,

2016; Hedge et al., 2021; Rey-Mermet et al., 2018, 2019; Rouder & Haaf, 2019; Rouder et al., 2019) and because we found in Draheim et al. that similar versions of the new and modified tasks included in the present study were vast improvements to these traditional tasks.

Except for the antisaccade task, the attention tasks included here were also modified from the Draheim et al. (2021) study to further improve their validity for assessing individual differences in attention control. Specifically, four separate selective visual arrays tasks were administered, SACT had its distractor removed and the visual cue no longer remained on the display during the wait time interval, and flanker DL and Stroop DL were made to be adaptive on a trial-by-trial basis instead of block-by-block as with the original study. Note that SACT and the visual arrays tasks had experimental manipulations as part of other projects, discussion of which goes beyond the scope of the present study.

Antisaccade

Antisaccade was unchanged from Draheim et al. (2021) as the version used here has consistently been shown to be a strongly reliable and valid measure of attention control across many of our studies (e.g., Draheim et al., 2021; Kane et al., 2001; Unsworth et al., 2004).

Color- and location-selective visual arrays

The visual arrays paradigm was created by Luck and Vogel (1997) to investigate the capacity of visual working memory and it is largely considered to be a working memory capacity task. However, recent evidence has challenged this view, particularly in versions of the task that include distractor stimuli (i.e., selective visual arrays). For example, several behavioral and electroencephalography studies have supported the strong role of attention in performing well on the task, such as that capacity scores strongly decline when set-sizes become larger than one's capacity (Fukuda et al., 2016), that the magnitude of event-related potentials associated with processing of information is similar in high- and low-cognitive ability individuals at smaller set sizes when there are not distractors present, but different when distractors *are* present (Vogel et al., 2005), and that visual arrays performance is at least partly a function of how long it takes individuals to recover from attentional capture (Fukuda & Vogel, 2011). As a result, while selective visual arrays tasks clearly involve a primary memory (i.e., short-term memory storage) component, *individual differences* in performance on them are more so driven by attentional factors than memory ones. In support of this, we have found throughout multiple datasets that selective visual arrays tasks: (a) correlate more strongly with other attention measures (namely antisaccade) than working memory capacity tasks, even though these attention control tasks have minimal storage demands, and (b) load more strongly onto factors with other attention tasks

than factors with working memory tasks in both exploratory and confirmatory factor analyses (see Draheim et al., 2021; Martin et al., 2021). And while visual arrays has good face validity as a measure of working memory capacity, closer examination of the demands of the task reveal that performance in selective versions of the task are largely driven by attentional factors. Namely, that performing well on the task requires the respondent to intensely apply attentional resources at a critical moment of the trial (similar to antisaccade) such that any micro- or macro-level attentional lapses will tank performance on the task, *and* the respondent must selectively attend to the only target stimuli while suppressing or otherwise ignoring the distractors.

The color-selection visual arrays task (color VA) used in the present study differed from the one employed on Draheim et al. (2021) in two ways. First, set sizes were reduced to 3 and 5 instead of 5 and 7. The rationale for this was partially based on unreported analyses from Draheim et al. (2021) that set-size 5 performance was slightly more predictive than set-size 7 performance, suggesting that larger set sizes are not necessary for the task to be a strong measure of attention control. Further, set size 3 was used to test whether set sizes that are around or even slightly before the capacity for the typical participant would still reflect attentional processing. In other words, reducing the primary memory demands of the task affect its strength as an attention control measure. It was found that set-size 3 and set-size 5 performance was roughly equivalent in terms of reliability and correlations to other tasks, and so in the present study performance on the two set sizes were aggregated in scoring. The other change to the task was that the number of target stimuli in the task (i.e., the set size) was identical to the number of distractors. This was incorrectly reported to be the case in Draheim et al. (p. 7) but during programming for the present study-at-large we realized that the version used in Draheim et al. always included 12 total stimuli such that five were targets and seven distractors in the set-size five trials, and seven were targets and five were distractors in the set-size seven trials.

The concentric visual arrays task (concentric VA) used in the present study was a novel task. It was included in the present study in part to test whether a visual arrays task with location selection could be as strong of an attention control measure as the color-selective version we have more commonly used. Location-selective visual arrays tasks are commonly used, but they typically involve presenting stimuli in different quadrants or hemifields, such that the respondent needs to look to a specific side of the display to catch the target stimuli (e.g., Shipstead et al., 2014). Performance on these tasks tends to be better than color-selection versions, suggesting it is easier to filter out distractors

based on location than color (e.g., Anllo-Vento & Hillyard, 1996; Vogel et al., 2005). The version used here was hypothesized to be *more* difficult than color selection by requiring participants to either constrain their visual attention spotlight (on trials with focal target stimuli) or diffuse their visual attentional spotlight in a donut-like shape (on trials with peripheral target stimuli). This donut-shaped allocation of visual attention is only sparsely referred to or studied in the literature. In an electroencephalography study, Müller and Hübner (2002) found that donut-shaped allocation seems to be substantively different from a spotlight-like allocation, and Beck and Lavie (2005) reported that when participants were asked to fixate on a location which will contain a distractor, the size of flanker interference effects were magnified. And while we did find that trials requiring spotlight allocation of visual attention were differentially predictive than trials requiring donut-shaped allocation of visual attention, the distinction between these types of trials goes beyond the scope of the paper and the differences were not so pronounced as to preclude aggregating these two trial types into one score for the task included in the present study.

Sustained attention-to-cue task (SACT)

SACT was a novel task included in Draheim et al. (2021) as an accuracy-based analog to the psychomotor vigilance task, which is a sustained attention task requiring participants to monitor a row of zeros and respond as quickly as they notice that these zeros have begun counting up. In Draheim et al., this task was both reliable and strongly inter-correlated with the other attention control measures, such as antisaccade and selective visual arrays. If this task measures the ability to sustain attention over the wait time interval, then it would be expected that performance at longer wait times should be worse compared to shorter wait times. However, this was not the case in Draheim et al., as performance was worse at the shortest wait time (2 s) and there were no differences in performance across the other wait times (4, 8, and 12 s). Further, in the original SACT, the visual cue remained on the display during the entire wait time interval and had an element very similar to antisaccade such that a distracting asterisk appeared away from the to-be-attended to location just prior to the onset of the target stimulus, thus potentially resulting in an artificially strong correlation to the antisaccade task. The visual cue remaining on the display may have allowed failures to sustain attention to occur and be corrected or using the distractor as a cue to shift attention back to the to-be-attended location without hurting performance on the task. Therefore, in this updated version of this task, the visual cue no longer remained on the display and the distracting stimulus was removed. The results of these

changes to performance on the task at different wait time intervals are beyond the scope of this paper and will be reported elsewhere in more detail. However, performance on this task did decrease with longer wait times—which is what we would expect if it measures an aspect of sustained attention over the wait time interval.

Flanker and Stroop tasks with adaptive response deadlines (flanker DL & Stroop DL)

Flanker and Stroop tasks are conflict-resolution tasks widely considered to require attention control in the form of cognitive inhibition and they are very commonly used in individual differences studies of attention control and related abilities (e.g., Rouder & Haaf, 2019). However, they have been repeatedly shown to be psychometrically poor measures of attention/inhibitory control when used for individual differences purposes (see Draheim et al., 2019, 2021), which was a primary motivation for us creating new measures of attention control. The adaptive Stroop and flanker tasks in Draheim et al. (2021) were designed to potentially improve the psychometrics of these tasks by avoiding scoring via difference scores and to account for speed–accuracy interactions/individual differences in speed–accuracy tendencies by assessing how quickly participants could respond at a particular level of accuracy. The flanker and Stroop DL tasks in that study were clear psychometric improvements to their traditional and difference score-based counterparts in that performance in them were roughly 3x more reliable and more strongly correlated to other attention control and executive functioning measures. Critically, this was also shown to not be simply because they placed more demands on speeded responding (i.e., required greater processing speed), as the incremental predictive validity of these tasks was still strong when processing speed was removed using confirmatory factor analysis.

The versions of these tasks used in Draheim et al. (2021) were adaptive on a block-by-block basis such that participants performed eighteen blocks each with 18 trials, and if participants were correct on at least 15 trials (83.3% accuracy), then the task would get more difficult (less time to respond) on the subsequent block. The dependent variable was what the starting response deadline would have been on a hypothetical 19th block. While these tasks worked well, this was a rather crude method and had some limitations. First, it necessitated a tremendous number of trials (324), although unreported analyses showed that the tasks were still quite reliable and valid when using scores after only half the trials had been administered. Second, to maintain a desirable 2:1 ratio between congruent and incongruent trials, the task needed to converge on a high accuracy level to avoid the situation in which participants could respond incorrectly on all six of the incongruent trials and still "advance" to the next level (i.e., the task would get more difficult).

Even still, participants could respond correctly on all twelve congruent trials and at chance performance on the incongruent trials and still advance to the next level. Finally, the block structure of the task meant that the adaptive procedure only kicked in after each set of 18 trials were administered, which is unsatisfactory for a number of reasons, including that it took high-ability participants hundreds of trials to reach their actual level of performance and that scores in the tasks were less continuous than would be preferred. As such, the version of the tasks used in the present study was adaptive after each incongruent trial, which is a sounder method that permitted convergence on a lower level of accuracy (which is desirable in adaptive procedures; see Kaernbach, 1991), required fewer trials, and would ostensibly provide a cleaner and more precise estimate of each participant's ability level.

Working memory capacity

The assessment of working memory capacity in Draheim et al. (2021) involved three two-block and larger-set size versions complex span tasks (operation span, symmetry span, and rotation span). Although complex span tasks are widely used and considered to be highly valid measures of working memory capacity (e.g., Conway et al., 2005), assessing a construct with one type of paradigm is an obvious limitation as it likely does not constitute a representative sampling of the underlying ability. As such, the present study included two spatial complex span tasks (symmetry span and rotation span) and two verbal running span tasks (digits and letters).

Fluid intelligence

The assessment of fluid intelligence involved administering a subset of the Raven's Advanced Progressive Matrices, letter sets, and number series. This was unchanged from Draheim et al. (2021).

Processing speed

Processing speed measures were included to test discriminant validity, namely the possibility that any observed improvements in construct coherence and predictive validity of the attention control measures included here could be attributable to extraneous factors such as individual differences in processing speed or motivation. In other words, participants who process information more quickly and/or are more motivated will do better across all tasks, thus potentially resulting in artificially high correlations both among the attention control measures and between attention control and the other cognitive constructs. The view that attention tasks are strongly contaminated with processing speed variance is a popular one. For example, Rouder et al. (2022) recently argued that the antisaccade task more strongly indexes processing speed than attention control, Hedge et al. (2021) found that traditional Stroop and flanker tasks reflect processing speed and response cautiousness but

not the ability to resolve cognitive conflict, and Rey-Mermet et al. (2019) used multiple methods to account for processing speed in their assessment of attention control, including adding baseline and adaptive trials to their antisaccade task.

Processing speed measures were included in Draheim et al. (2021) for the same reason, but only to a subset of the final sample, as these tasks were included in a fifth session that was added to that study while data collection was already underway. In the present study, all participants performed the various processing speed measures, which were letter-string comparison, digit-string comparison,

and pattern comparison. The first two were also included in Draheim et al., but the pattern comparison was new to this study. It was included instead of the digit-symbol substitution task which is a more complicated processing speed task that places more attentional demands than a pure measure of processing speed.

Appendix B

All task-level correlations

Table 5 Full correlation matrix

	1.	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*	*13.*	*14.*	*15.*
1. Antisaccade															
2. SACT	**0.35**														
3. FlankerDL	− 0.08	− 0.09													
4. StroopDL	**− 0.35**	**− 0.24**	0.11												
5. VA – color	**0.41**	**0.35**	**− 0.12**	**− 0.28**											
6. VA – concentric	**0.42**	**0.41**	**− 0.13**	**− 0.36**	**0.72**										
7. SymSpan	**0.25**	**0.14**	0.04	**− 0.20**	**0.43**	**0.31**									
8. RotSpan	**0.27**	**0.16**	− 0.03	**− 0.19**	**0.34**	**0.34**	**0.56**								
9. RunLetter	**0.21**	**0.16**	− 0.02	**− 0.14**	**0.24**	**0.25**	**0.19**	**0.21**							
10. RunDigit	**0.25**	**0.20**	0.05	**− 0.23**	**0.23**	**0.23**	**0.22**	**0.20**	**0.64**						
11. RAPM	**0.33**	**0.12**	− 0.09	**− 0.14**	**0.44**	**0.40**	**0.33**	**0.33**	**0.29**	**0.26**					
12. Letter sets	**0.27**	0.10	0.04	**− 0.20**	**0.28**	**0.29**	**0.31**	**0.18**	**0.36**	**0.43**	**0.39**				
13. Number series	**0.28**	0.08	0.01	**− 0.26**	**0.42**	**0.36**	**0.33**	**0.24**	**0.25**	**0.29**	**0.46**	**0.57**			
14. Digit comp.	**0.33**	**0.20**	− 0.06	**− 0.28**	**0.33**	**0.29**	**0.24**	**0.22**	**0.18**	**0.29**	**0.26**	**0.43**	**0.38**		
15. Letter comp.	**0.22**	**0.17**	0.07	**− 0.21**	**0.25**	**0.18**	**0.22**	**0.19**	**0.21**	**0.24**	**0.18**	**0.41**	**0.29**	**0.60**	
16. Pattern comp.	**0.34**	**0.23**	0.10	**− 0.18**	**0.42**	**0.36**	**0.33**	**0.31**	0.07	**0.16**	**0.39**	**0.30**	**0.36**	**0.48**	**0.40**

Computed correlation used Pearson method with pairwise deletion. FlankerDL = Flanker with adaptive response deadline; StroopDL = Stroop with adaptive response deadline; VA = Visual arrays; SymSpan = Symmetry span; RotSpan = Rotation span; RunLetter = Running letter span; RunDigit = Running digit span; RAPM = Raven's advanced progressive matrices; Comp = Comparison. Values in boldface were statistically significant at α = .05
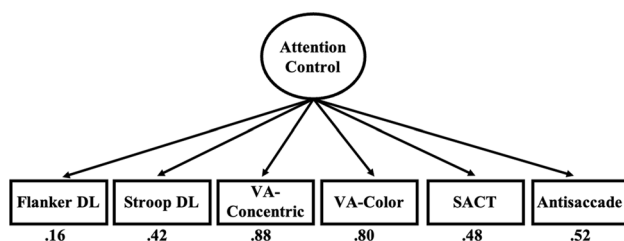
Appendix C

Additional analyses

Fig. 8 Confirmatory factor analysis with all attention control tasks
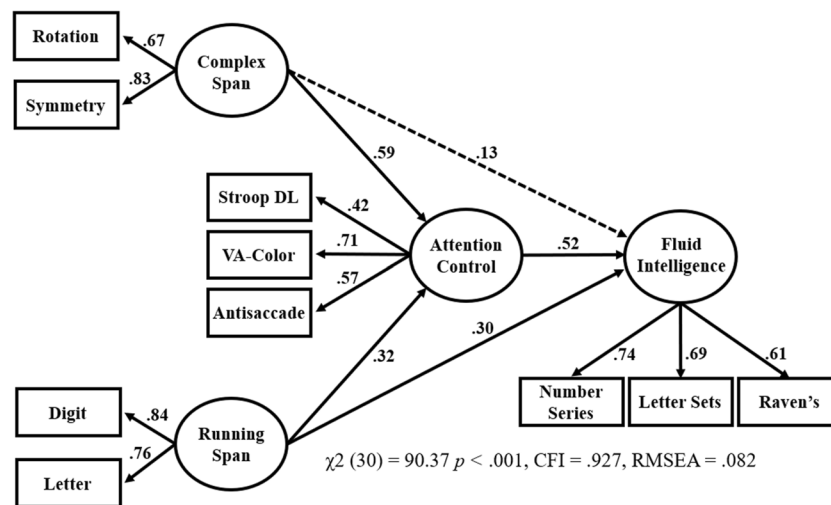
Fig. 9 Model showing full mediation of attention control on the working memory capacity–fluid intelligence relationship. *Note.* As in Draheim et al. (2021), some mediation models in the present study revealed that attention control fully mediated the relationship between complex span performance and fluid intelligence, depending on which tasks were included as indices of attention control. This is one such model – as the *dashed line* indicates a non-statistically significant path between fluid intelligence and complex span when accounting for attention control.

Declarations

Conflict of interest The authors have no conflicts of interest to disclose.

Ethics approval The Georgia Institute of Technology Institutional Review Board approved the protocols (H20165 & H20532) and amendments for this study throughout data collection.

Consent to participate Informed consent was obtained from all individual participants included in this study.

Consent to publish Participants consented to have their data disseminated in de-identifiable form.

References

Anllo-Vento, L., & Hillyard, S. A. (1996). Selective attention to the color and direction of moving stimuli: Electrophysiological correlates of hierarchical feature selection. *Perception & Psychophysics, 58*(2), 191–206. https://doi.org/10.3758/BF03211875

Beck, D. M., & Lavie, N. (2005). Look here but ignore what you see: effects of distractors at fixation. *Journal of Experimental Psychology: Human Perception and Performance, 31*(3), 592–607. https://doi.org/10.1037/0096-1523.31.3.592

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42*(2), 563–570. https://doi.org/10.3758/brm.42.2.563

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

Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., & Engle, R. W. (2022). Attention control and process overlap theory: Searching for cognitive processes underpinning the positive manifold. *Intelligence, 91*, 101629. https://doi.org/10.1016/j.intell.2022.101629

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.* Advance online publication. https://doi.org/10.1037/xge0001408

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, Z. D., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769–786. https://doi.org/10.3758/bf03196772

Conway, A. R. A., Kovacs, K., Hao, H., Rosales, K. P., & Snijder, J.-P. (2021). Individual differences in attention and intelligence: A united cognitive/psychometric approach. *Journal of Intelligence, 9*(3), 34. https://doi.org/10.3390/jintelligence9030034

Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining reaction time and accuracy: The relationship between working memory capacity and task-switching as a case example. *Perspectives on Psychological Science, 11*(1), 133–155. https://doi.org/10.1177/1745691615596990

Draheim, C., Harrison, T. L., Embretson, S. E., & Engle, R. W. (2018). What item response theory can tell us about the complex span tasks. *Psychological Assessment, 30*(1), 116–129. https://doi.org/10.1037/pas0000444

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

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

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

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

Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment, 8*(4), 341–349. https://doi.org/10.1037/1040-3590.8.4.341

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

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of atarget letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/BF03203267

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interferencecontrol functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Fukuda, K., & Vogel, E. K. (2009). Human variation in overriding attentional capture. *Journal of Neuroscience, 29*(27), 8726–8733. https://doi.org/10.1523/JNEUROSCI.2145-09.2009

Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentionalcapture. *Psychological Science, 22*(3), 361–368. https://doi.org/10.1177/0956797611398493

Fukuda, K., Woodman, G. F., & Vogel, E. K. (2016). Individual differences in visual workingmemory capacity: Contributions of attentional control to storage. In P. Jolicoeur, C. Lefebvre, & J. Martinez-Trujillo (Eds.), *Mechanisms of sensory working memory: Attention and performance XXV* (pp. 105–119). Elsevier Academic Press. https://doi.org/10.1016/b978-0-12-801371-7.00009-0

Gonthier, C. (2022). An easy way to improve scoring of memory span tasks: The edit distance, beyond "correct recall in the correct serial position". *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-022-01908-2

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

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hedge, C., Powell, G., Bompas, A., & Sumner, P. (2021). Strategy and processing speed eclipseindividual differences in control ability in conflict tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0001028

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 645–662. https://doi.org/10.1037/0278-7393.33.4.645

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

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General, 130*(2), 169–183. https://doi.org/10.1037//0096-3445.130.2.169

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281. https://doi.org/10.1038/36846

Martin, J. D., Tsukahara, J. S., Draheim, C., Shipstead, Z., Mashburn, C. A., Vogel, E. K., & Engle, R. W. (2021). The Visual Arrays Task: Visual Storage Capacity or AttentionControl?. *Journal of Experimental Psychology: General*. Advance online publication. https://doi.org/10.1037/xge0001048

Müller, M. M., & Hübner, R. (2002). Can the spotlight of attention be shaped like a doughnut? Evidence from steady-state visual evoked potentials. *Psychological Science, 13*(2), 119–124. https://doi.org/10.1111/1467-9280.00422

Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review, 20*(5), 819–858. https://doi.org/10.3758/s13423-013-0404-5

Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory? An examination of the relationships among working memory, secondary memory, and fluid intelligence. *Psychological Science, 19*(11), 1071–1077. https://doi.org/10.1111/j.1467-9280.2008.02202.x

Nieuwenhuis, S., Stins, J. F., Posthuma, D., Polderman, T. J., Boomsma, D. I., & de Geus, E. J. (2006). Accounting for sequential trial effects in the flanker task: Conflict adaptation or associative priming? *Memory & Cognition, 34*(6), 1260–1272. https://doi.org/10.3758/bf03193270

Norman, D. A., & Shallice, T. (1986). Attention to action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation*. Springer. https://doi.org/10.1007/978-1-4757-0629-1_1

Paap, K. R., & Sawi, O. (2016). The role of test–retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods, 274*, 81–93. https://doi.org/10.1016/j.jneumeth.2016.10.002

Pollack, I., Johnson, I. B., & Knaff, P. R. (1959). Running memory span. *Journal ofExperimental. Psychology, 57*(3), 137–146. https://doi.org/10.1037/h0046137

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing https://www.R-project.org/

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In R. S. McCallum (Ed.), *Handbook of Nonverbal Assessment*. Springer. https://doi.org/10.1007/978-1-4615-0153-4_11

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin and Review, 20*(6), 1102–1113. https://doi.org/10.3758/s13423-013-0453-9

Redick, T. S., Unsworth, N., Kelly, A. J., & Engle, R. W. (2012). Faster, smarter? Working memory capacity and perceptual speed in relation to fluid intelligence. *Journal of Cognitive Psychology, 24*(7), 844–854. https://doi.org/10.1080/20445911.2012.704359

Revelle, W. (2022). *psych: Procedures for psychological, psychometric, and personality research* [Manual]. https://CRAN.R-project.org/package=psych

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 501–526. https://doi.org/10.1037/xlm0000450

Rey-Mermet, A., Gade, M., Souza, A. S., Von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General, 148*(8), 1335–1372. https://doi.org/10.1037/xge0000593

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

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

Rouder, J., Kumar, A., & Haaf, J. M. (2019). Why Most Studies of Individual Differences With Inhibition Tasks Are Bound To Fail. *PsyArXiv.* https://doi.org/10.31234/osf.io/3cjr5

Rouder, J. N., de la Pena, A. C., Pratte, M., Richards, V., Hernan, M., Pascoe, M., & Thapar, A. (2022). Is the Antisaccade Task a Unicorn Task For Measuring Cognitive Control? *PsyArXiv.* https://doi.org/10.31219/osf.io/fhg3n

Rueda, M. R. (2018). Attention in the heart of intelligence. *Trends in Neuroscience and Education, 13*, 26–33. https://doi.org/10.1016/j.tine.2018.11.003

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

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory, 20*(6), 608–628. https://doi.org/10.1080/09658211.2012.691519

Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language, 72*(1), 116–141. https://doi.org/10.1016/j.jml.2014.01.004

Shipstead, Z., Harrison, T. L., & Engle. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science, 11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Stoffels, E. J., & Van der Molen, M. W. (1988). Effects of visual and auditory noise on visual choice reaction time in a continuous-flow paradigm. *Perception & Psychophysics, 44*(1), 7–14. https://doi.org/10.3758/bf03207468

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Thurstone, L. L. (1938). *Primary Mental Abilities*. University of Chicago Press.

Tsukahara, J.S. (2022). englelab: An R package for processing complex-span and attention control tasks downloaded from the EngleLab (1.1.0). Zenodo. https://doi.org/10.5281/zenodo.6987145

Tsukahara, J. S., Harrison, T. L., Draheim, C., Martin, J. D., & Engle, R. W. (2020). Attention control: The missing link between sensory discrimination and intelligence. *Attention, Perception, and Psychophysics, 82*(7), 3445–3478. https://doi.org/10.3758/s13414-020-02044-9

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*(4), 392–406. https://doi.org/10.1016/j.jml.2010.02.001

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1302–1321. https://doi.org/10.1037/0278-7393.30.6.1302

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505. https://doi.org/10.3758/bf03192720

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*(6), 635–654. https://doi.org/10.1080/09658210902998047

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

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: a meta-analysis. *Psychology and Aging, 13*(1), 120–126. https://doi.org/10.1037/0882-7974.13.1.120

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature, 438*(7067), 500–503. https://doi.org/10.1038/nature04171

Von Bastian, C. C., Blais, C., Brewer, G. A., Gyurkovics, M., Hedge, C., Kalamala, P., Meier, M. E., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, A. S., Conway, A. R. A., Draheim, C., Engle, R. W., Fischkorn, G. T., Friedman, N. P., Gustavson, D. E., Koch, I., et al. (2020). Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations. *PsyArXiv.* https://doi.org/10.31234/osf.io/x3b9k

Wickelgren, W. A. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41*(1), 67–85. https://doi.org/10.1016/0001-6918(77)90012-9

Zheng, Y., Myerson, J., & Hale, S. (2000). Age and individual differences in visuospatial processing speed: Testing the magnification hypothesis. *Psychonomic Bulletin & Review, 7*, 113–120. https://doi.org/10.3758/BF03210729