Is executive control related to working memory capacity and fluid intelligence

Alodie Rey-Mermet, Miriam Gade, Alessandra S. Souza, Claudia C. von Bastian ...+1 more authors

Institutions: University of Zurich, University of Sheffield

Published on: 08 Apr 2019 - Journal of Experimental Psychology: General (American Psychological Association)

Topics: Working memory

Year: 2019

Is executive control related to working memory capacity and fluid intelligence?

Rey-Mermet, Alodie ; Gade, Miriam ; Souza, Alessandra S ; Von Bastian, Claudia C ; Oberauer, Klaus

Abstract: In the last two decades, individual-differences research has put forward 3 cognitive psychometric constructs: executive control (i.e., the ability to monitor and control ongoing thoughts and actions), working memory capacity (WMC, i.e., the ability to retain access to a limited amount of information in the service of complex tasks), and fluid intelligence (gF, i.e., the ability to reason with novel information). These constructs have been proposed to be closely related, but previous research failed to substantiate a strong correlation between executive control and the other two constructs. This might arise from the difficulty in establishing executive control as a latent variable and from differences in the way the 3 constructs are measured (i.e., executive control is typically measured through reaction times, whereas WMC and gF are measured through accuracy). The purpose of the present study was to overcome these difficulties by measuring executive control through accuracy. Despite good reliabilities of all measures, structural equation modeling identified no coherent factor of executive control. Furthermore, WMC and gF-modeled as distinct but correlated factors-were unrelated to the individual measures of executive control. Hence, measuring executive control through accuracy did not overcome the difficulties of establishing executive control as a latent variable. These findings call into question the existence of executive control as a psychometric construct and the assumption that WMC and gF are closely related to the ability to control ongoing thoughts and actions. (PsycINFO Database Record (c) 2019 APA, all rights reserved).

DOI: https://doi.org/10.1037/xge0000593

Originally published at:
Rey-Mermet, Alodie; Gade, Miriam; Souza, Alessandra S; Von Bastian, Claudia C; Oberauer, Klaus (2019). Is executive control related to working memory capacity and fluid intelligence? Journal of Experimental Psychology: General, 148(8):1335-1372.
DOI: https://doi.org/10.1037/xge0000593

Is executive control related to working memory capacity and fluid intelligence?

Alodie Rey-Mermet [1, 2]

Miriam Gade [1, 2. 4]

Alessandra Souza [1]

Claudia C. von Bastian [3]

Klaus Oberauer [1]

[1] University of Zurich and University Research Priority Program „Dynamics of Healthy Aging"

[2] Catholic University of Eichstätt-Ingolstadt

[3] University of Sheffield

[4] Medical School Berlin

Author Note

Alodie Rey-Mermet, Department of Psychology, Cognitive Psychology Unit, University of Zurich, Switzerland, University Research Priority Program (URPP) "Dynamics of Healthy Aging", University of Zurich, Switzerland, and Department of Psychology, General Psychology, Catholic University of Eichstätt-Ingolstadt, Germany; Miriam Gade, Department of Psychology, Cognitive Psychology Unit, University of Zurich, Switzerland, University Research Priority Program (URPP) "Dynamics of Healthy Aging", University of Zurich, Switzerland, and

Department of Psychology, General Psychology, Catholic University of Eichstätt-Ingolstadt, Germany; Medical School Berlin, Germany; Alessandra Souza, Department of Psychology, Cognitive Psychology Unit, University of Zurich, Switzerland, and University Research Priority Program (URPP) "Dynamics of Healthy Aging", University of Zurich, Switzerland; Claudia C. von Bastian, Department of Psychology, University of Sheffield, United Kingdom; Klaus Oberauer, Cognitive Psychology Unit, Department of Psychology, University of Zurich, Switzerland.

This work was financially supported by the University Research Priority Program „Dynamics of Healthy Aging" of the University of Zurich and by a grant from the Biäsch Foundation to M. Gade, A. Souza and C. von Bastian. Some results of this article were presented at the 59th Conference of Experimental Psychologists (TeaP), Dresden, Germany, March 2017, at the 20th Conference of the European Society for Cognitive Psychology (ESCoP), Potsdam, Germany, September 2017, and at the 58th Annual Meeting of the Psychonomic Society, Vancouver, Canada, November 2017. We thank Dave Koller, Julia Moser, and Sarah Schoch for their help in data collection. The test battery, data, scripts analyses and results presented in this work can be found at https://osf.io/4t4h5.

Correspondence should be addressed to Alodie Rey-Mermet, Catholic University of Eichstätt-Ingolstadt, Department of Psychology, General Psychology, Ostenstr. 27, 85072 Eichstätt, email alodie.rey-mermet@ku.de, phone: +49 8421 93 23232, fax: +49 8421 93 21074

Word count: 13'373

Abstract

In the last two decades, individual differences research put forward three cognitive psychometric constructs: executive control (i.e., the ability to monitor and control ongoing thoughts and actions), working memory capacity (WMC, i.e., the ability to retain access to a limited amount of information in the service of complex tasks) and fluid intelligence (gF, i.e., the ability to reason with novel information). These constructs have been proposed to be closely related, but previous research failed to substantiate a strong correlation between executive control and the other two constructs. This might arise from the difficulty in establishing executive control as a latent variable and from differences in the way the three constructs are measured (i.e., executive control is typically measured through reaction times, whereas WMC and gF are measured through accuracy). The purpose of the present study was to overcome these difficulties by measuring executive control through accuracy. Despite good reliabilities of all measures, structural equation modeling identified no coherent factor of executive control. Furthermore, WMC and gF – modeled as distinct but correlated factors – were unrelated to the individual measures of executive control. Hence, measuring executive control through accuracy did not overcome the difficulties of establishing executive control as a latent variable. These findings call into question the existence of executive control as a psychometric construct and the assumption that WMC and gF are closely related to the ability to control ongoing thoughts and actions.

Keywords: executive functions, cognitive control, attentional control, individual differences

Is executive control related to working memory capacity and fluid intelligence?

Executive control – also referred to as cognitive control, attention/attentional control, executive attention or executive functions – is the ability to supervise and control thoughts and actions in order to achieve current goals (e.g., Burgess, 1997). In individual differences research, executive control has been put forward as explaining substantial variability in both working memory capacity (WMC) and fluid intelligence (gF; Engle, 2002; Engle & Kane, 2004; Kane, Conway, Hambrick, & Engle, 2007). WMC is the ability to retain access to a limited amount of information in the service of complex tasks (e.g., Baddeley & Hitch, 1974; Cowan, 1998; Miyake & Shah, 1999; Oberauer, 2009; Oberauer & Kliegl, 2006), and gF is the ability to reason with novel information (e.g., Cattell, 1963). Although some recent attempts confirmed a strong correlation between executive control and the two other constructs (e.g., Shipstead, Lindsey, Marshall, & Engle, 2014; Unsworth, Spillers, & Brewer, 2009), others failed to do so (e.g., Chuderski, Taraday, Nęcka, & Smoleń, 2012; Friedman et al., 2006). One reason for this mixed evidence might be the difficulty of establishing executive control as a coherent psychometric construct (e.g., Karr et al., 2018; Rey-Mermet, Gade, & Oberauer, 2018). Another reason might be that executive control is typically measured through the speed of processing as indexed by reaction times (RTs), whereas WMC and gF are measured through the accuracy of processing (i.e., the proportion of correct responses). The purpose of the present study was to establish executive control as a psychometric construct and to determine to what extent executive control is related to WMC and gF when all three constructs are measured through the accuracy achieved within a limited amount of time.

Measuring Individual Differences in Executive Control, WMC, and gF

Executive control has been argued to be the primary source of individual differences in

WMC and gF (e.g., Engle, 2002; Engle & Kane, 2004; Kane, Bleckley, Conway, & Engle, 2001; Kane et al., 2007). To support this claim, early studies started from the conceptualization of WMC as a composite of short-term memory (STM) and executive control. For instance, Engle et al. (1999) measured executive control as the residual variance of WMC after controlling for short-term memory. In Kane et al. (2004), executive control was modeled as the common variance shared by all WMC and STM tasks. In both studies, this latent measure of executive control was found to be correlated with gF (.52 and .49, respectively). The conclusions from these studies hinge critically on the assumption that the variance component they isolated from WMC task scores reflects executive control. As Engle and colleagues admit, this "is the result of a logical analysis [but] is, at best, an educated conjecture" (Engle et al., 1999, p. 326).

To address the role of the executive-control construct in the relationship between WMC and gF, subsequent research involved tasks assumed to assess executive control more directly (see Table 1 for a summary of their findings, and Table A1 in the Appendix A for a description of the tasks). At first glance, this research has revealed predominantly positive evidence for the relation of executive control with WMC and gF (see Table 1). A closer look, however, reveals four prevalent issues in most studies reported in Table 1: (1) the difficulty to establish a coherent factor of executive control, (2) the predominance of one executive-control task's variance if a factor was established, (3) the confounding of executive control and general processing speed, and (4) the mismatch in how executive control, WMC and gF are measured. We next discuss each of these problems in detail.

Failure to Establish a Coherent Factor of Executive Control

A first prerequisite for testing the assumption that executive control is related to WMC and gF is to establish the construct validity of executive control. That means that multiple measures of executive control correlate substantially with each other, so that their shared variance can be represented by a latent variable (i.e., a factor), or a set of interrelated factors. In their seminal work, Miyake et al. (2000) have reported a model in which executive control was represented by three factors: inhibition (i.e., the ability to ignore and suppress irrelevant ongoing thoughts and actions), updating (i.e., the replacement of the currently stored material in WM by new information), and task-switching (i.e., the ability to shift the attention to other tasks or perceptual dimension). Although their model has been replicated (Friedman et al., 2006), there are, at the same time, several reports questioning the validity of the executive-control construct (see Karr et al., 2018), in particular the validity of the inhibition construct. Whereas several studies found positive correlations between at least some indicators of inhibition (e.g., Chuderski et al., 2012, Exp. 1; Friedman & Miyake, 2004; Kane et al., 2016; Pettigrew & Martin, 2014; Stahl et al., 2014), many studies also reported low zero-order correlations between inhibition tasks (e.g., De Simoni & von Bastian, 2017; Guye & von Bastian, 2017; Paap & Greenberg, 2013; Rey-Mermet et al., 2018; von Bastian, Souza, & Gade, 2016). In some studies, the tasks assumed to measure inhibition did not load on a coherent latent variable (Brydges, Reid, Fox, & Anderson, 2012; Hull, Martin, Beier, Lane, & Hamilton, 2008; Krumm et al., 2009; Rey-Mermet et al., 2018). In other studies, inhibition measures had to be merged with tasks assumed to measure general processing speed to create a coherent latent variable (Hedden & Yoon, 2006; van der Sluis, de Jong, & van der Leij, 2007).

Why is it so difficult to establish executive control, and in particular inhibition, as a coherent latent variable? One reason might be that executive control is mainly assessed through RTs, and different participants can have different speed-accuracy tradeoffs. That is, some participants can favor accuracy over speed, whereas others do the reverse by favoring speed over accuracy. Thus, when the focus is on the RTs only, the variance of accuracy is neglected. One approach to solve this problem would be to combine both dependent measures (i.e., RTs and accuracy) into a single measure. There is, however, no principled way of combining separate measures of RT and accuracy into a single score (Bruyer & Brysbaert, 2011; Dennis & Evans, 1996; Hughes, Linck, Bowles, Koeth, & Bunting, 2014; Vandierendonck, 2017).

Predominance of Single Tasks

A closer examination of those studies that did establish an executive-control factor (see Table 1) reveals that this factor was frequently dominated by one task with a high loading, whereas the other tasks had very low loadings and high error variances (Chuderski et al., 2012; Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010, Exp. 2; McVay & Kane, 2012; Shipstead et al., 2014; Unsworth, 2015, Exp. 3; Unsworth & McMillan, 2014; Unsworth & Spillers, 2010). This implies that the executive-control factor in these studies does not represent much common variance among multiple measures of executive control. Thus, in many studies, the executive-control factor was mainly driven by one task, which may or may not be related to WMC or gF (see Table B1 in Appendix B for an overview of the factor loadings). As shown in Table 1, in most of these cases, the task with the highest loading was the antisaccade task, in which participants need to quickly move their eyes in the direction opposite of a peripheral cue to identify a briefly presented stimulus.

Confound of Executive-Control Measures with General Processing Speed

Prior studies that successfully established a latent executive-control factor typically used RT differences as dependent measure (e.g., Klauer et al., 2010; Shipstead et al., 2014; Unsworth, Spillers, et al., 2009). For example, in the color Stroop task, participants are asked to indicate the print color of a color word while ignoring the meaning of the word. To this end, participants encounter incongruent trials (i.e., trials with some form of conflict between relevant and irrelevant information, such as the word “green” printed in red), congruent trials (i.e., trials without conflict between response features, such as the word “red” printed in “red”) and/or neutral trials (i.e., trials with only one response-relevant feature, such as "xxxxx" printed in red). Executive control is required in incongruent trials to ignore or suppress the irrelevant information that creates a conflict (i.e., the meaning of the word).

To isolate the executive-control processes from non-executive-control processes, RTs on baseline trials (i.e., congruent or neutral trials) are typically subtracted from RTs on incongruent trials (e.g., Miyake et al., 2000; Unsworth, Spillers, et al., 2009). Subtracting RTs is premised on the assumption of additive factors. That is, the duration of the processes in the baseline condition and the duration of the executive-control process combine additively to the RT in incongruent trials, and the duration of each process is uniquely affected by its own source of individual differences. However, this assumption is questionable because across various speeded tasks, the RTs of slow individuals are related to those of fast individuals through a constant proportional slowing factor (Zheng, Myerson, & Hale, 2000). This implies that differences between RTs are also proportionally larger for slower than for faster individuals. For instance, a generally faster person may have RTs of 400 and 600 ms in the baseline and incongruent trials, respectively, while a generally slower person would have RTs of 600 and 900 ms, respectively. In both cases, the two RTs are related to each other by the same proportional increase (by a factor of 1.5). Subtracting RTs would, however, result in smaller interference scores in the first relative to the second person (i.e., 200 and 300 ms, respectively), thus creating differences between individuals. Thus, these individual differences in the RT-difference score are caused entirely by differences in general processing speed, suggesting that these measures of executive control were likely to be confounded with general processing speed (see Jewsbury, Bowden, & Strauss, 2016). This raises the possibility that the studies that did establish a factor of executive control (see Table 1), have reported shared variance in processing speed rather than in executive control. As general processing speed has been found to correlate with WMC and gF (e.g., Keye, Wilhelm, Oberauer, & Ravenzwaaij, 2009; Keye, Wilhelm, Oberauer, & Stürmer, 2013; Krumm et al., 2009; Wilhelm & Oberauer, 2006; but see Conway, Cowan, Bunting, Therriault, & Minkoff, 2002), this could (at least, partially) explain the observed correlations between executive control and gF or WMC.

Mismatch of Method Variance Between Executive Control and WMC or gF

The presence or absence of a link between executive control and WMC or gF could also be driven by a mismatch of method variance, with executive control being measured through RT, and WMC and gF typically being measured through the accuracy achieved within the available processing time. Notably, the antisaccade task is one of the few tasks measuring executive control through accuracy so that using it results in a partial match of method variance. Hence, this could explain why substantial correlations between executive control and the two other constructs were found in many studies in which the antisaccade task played a dominant role in measuring executive control (see Table 1).

The Present Study

The present study had two purposes. The first goal was to determine whether executive control could be established as a psychometric construct. The second purpose was to examine the relationship between executive control and the other two constructs, that is, WMC and gF. To this end, we measured executive control through accuracy within limited available time, allowing for measuring executive control, gF and WMC on the same scale. To do that, we used a calibration procedure that adjusted a response deadline limiting the time participants had to respond. Our rationale was as follows: When measured with the conventional self-paced method, poor executive control can translate into slower RTs, more errors, or both, in those conditions that require more executive control than in the baseline condition (e.g., incongruent trials vs. congruent trials). With a response deadline, both responses that are too slow and erroneous responses translate into a reduction of accuracy. That is, if a person with relatively poor executive control favors accuracy over speed in the response deadline procedure, s/he would miss the deadline more often on incongruent trials than on congruent trials. In contrast, if a person with relatively low executive-control ability favors speed over accuracy, s/he would select the wrong response more often on incongruent trials than on congruent trials. Thus, it does not matter whether a person fails more often on incongruent than congruent trials because they miss the deadline or because they select the wrong response; inefficient executive control would be translated into lower accuracy. In this way, individual differences in executive control are completely mapped onto a single measurement scale, namely accuracy within the available time, irrespective of the speed-accuracy trade-offs.

We calibrated the duration of the response deadline for each participant to achieve a fixed accuracy level in blocks in which neutral trials were presented. In subsequent experimental blocks, incongruent and congruent trials were presented, and the response deadline was fixed to the calibrated duration. We reasoned that accurate responding to incongruent trials should require more time than responding to baseline trials, but because of the deadline, this increased time demand on executive control should be translated into lower accuracy. We see several advantages in using the calibration procedure. First, the calibration procedure reduces individual differences in general ability and in the ability to carry out all the processes involved in the neutral trials that also contribute to performance in incongruent trials. Second, it also removes individual differences in speed-accuracy tradeoffs, because, irrespective of whether a participant favored accuracy over speed or speed over accuracy, inefficient executive control would result in lower accuracy. Third, calibration is superior to subtracting RTs because calibration moves every individual into roughly the same point on the measurement scale. However, calibration may still involve some unsystematic variance if it does not move every individual to the exact same point of the measurement scale. To remove this noisiness, we used the difference in accuracy rates between incongruent and congruent trials as measure of executive control.

In the present study, we focused on inhibition tasks to measure executive control for three reasons. First, most previous research has primarily measured executive control through tasks assumed to assess inhibition (see Table A1 in the Appendix A). Second, in their updated model, Miyake and Friedman (Friedman & Miyake, 2016; Miyake & Friedman, 2012) assumed that the tasks used to assess inhibition involved the basic ability necessary for all three executive functions (i.e., inhibition, updating, and task-switching; see also Munakata et al., 2011). Third, measuring inhibition as a coherent latent variable seems more challenging (Rey-Mermet et al., 2018) than measuring task-switching (von Bastian & Druey, 2017) or updating of WM contents (Ecker, Lewandowsky, Oberauer, & Chee, 2010). For example, in a previous study (Rey-Mermet et al., 2018), we aimed to determine the psychometric structure of inhibition among a large set of tasks widely used in the existing literature to measure this construct. The dependent measures – except for the antisaccade task – were based on RTs. Although a factorial measurement model was identified, this model had low explanatory power as each factor was dominated by one task with a high loading, whereas the other tasks had very low loadings and high error variances.

In the present study, we hypothesized that if the difficulty to establish a latent variable of executive control stems from the neglect of accuracy variance and/or from subtracting RTs, we should be able to find a latent variable representing executive control because executive control was measured through accuracy. Moreover, in this case, if the mixed evidence regarding the correlations between executive control and WMC or gF in previous studies (see Table 1) stems from the difference in measurement scale (i.e., RT vs. accuracy), we would expect to find substantial correlations between all three constructs in the present study. However, in the case that executive control could not be identified as a latent variable, we planned to investigate the relations between the individual measures of executive control and the gF and WMC constructs.

Method

Across both Method and Results sections, we report how we determined our sample size and all data exclusions, manipulations, and measures in the study (Simmons, Nelson, & Simonsohn, 2012).

Participants

We aimed at a sample size of 160-200, which we determined based on previous research (e.g., Chuderski et al., 2012; Unsworth, Fukuda, Awh, & Vogel, 2014; Unsworth & Spillers, 2010). Students from Swiss universities (University of Zurich, ETH Zurich) were recruited. In total, 196 participants were tested. All reported Swiss German or German as native language, normal or corrected-to-normal vision, and no color blindness. Fifteen participants had missing data (eight in executive-control tasks, two in WMC tasks, one in half of the tasks because the computer malfunctioned, and four participants in the time estimation task – see below – because they used the wrong key to respond). The final sample consisted of 181 participants. Demographic characteristics are summarized in Table 2.

The study was carried out according to the guidelines of the ethics committee of the

Faculty of Arts and Social Sciences at the University of Zurich, and all participants gave written informed consent. At study completion, participants received either course credits or CHF 60 (about USD 60).

Task Materials

For each construct (i.e., executive control, WMC or gF), we opted for tasks for which there is broad (though not necessarily universal) agreement that the relevant process plays a role for successful performance. Moreover, we included tasks typically used in previous individual differences research (e.g., Friedman & Miyake, 2004; Kane et al., 2004, 2016; Miyake et al., 2000; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Rey-Mermet et al., 2018; Shipstead et al., 2014; von Bastian et al., 2016). Hence, seven tasks were used to measure executive control (i.e., the color Stroop, number Stroop, arrow flanker, letter flanker, Simon, antisaccade, and stop-signal tasks). Five tasks were used to measure gF (i.e., the locations test, the letter sets test, the Raven's advanced progressive matrices test, the nonsense syllogisms test, and the diagramming relationships test). Four tasks were used to measure WMC (i.e., an updating and a complex span task, each with numerical and spatial materials). Finally, a time estimation task was also used to measure how good participants could estimate the deadline presented in form of a visual timer.

All tasks were programmed using Tatool (von Bastian, Locher, & Ruflin, 2013), and run on IBM compatible computers. For each task, the same pseudorandom trial sequence was administered for all participants. The number of trials and of blocks are summarized in Appendix C.

Executive-control tasks. As executive-control measures may be affected by episodic memory and associative learning (e.g., Hommel, 2004; Mayr, Awh, & Laurey, 2003), we applied three constraints to reduce the impact of memory contributions in each executive-control task.

First, the trial sequence did not contain any trial-to-trial repetitions of the exact stimulus, reducing the influence of trial-to-trial episodic memory. Second, trial types (e.g., incongruent and congruent trials), response keys, presentation location, and individual stimulus exemplars were counterbalanced across trials as far as possible to minimize associative learning. Third, stimulus material did not overlap across tasks to minimize carry-over effects of any learning between tasks. Hereafter, we describe the calibration procedure we used for most tasks used to assess executive control (i.e., all tasks, except the stop-signal task). Then, we describe each task in details.

Calibration Procedure for the Response Deadline. The response deadline was calibrated individually according to an adaptive rule, the weighted up-down method (Kaernbach, 1991). That is, each correct response led to a decrease in response deadline, and each incorrect response to an increase. For most tasks (i.e., the color Stroop, number Stroop, arrow flanker, letter flanker and Simon tasks), we calibrated the response deadline to achieve 75% correct responses. As three tasks – that is, the arrow flanker, letter flanker and Simon tasks – were two-choices tasks (with 50% being the chance level for two-choices tasks), 75% as cut-off yields the same variability range for performance above and below the cut-off. Placing our measures halfway between floor and ceiling on the measurement scale maximizes the sensitivity of the measurement scale. Moreover, although the color Stroop and number Stroop tasks were four-choice tasks, we used the same cut-off of 75% for these tasks because pilot studies showed that with lower cut-offs accuracy did not converge well to the criterion.

In order to converge to 75% of correct responses, the response deadline was adapted stepwise using larger adjustments for upward steps (i.e., increases in presentation times) and smaller adjustments for downward steps (i.e., decreases in presentation times). As the monitors had a refresh rate of 60 Hz (i.e., the display was refreshed 60 times per second), each refresh cycle was 16.67 ms (which was calculated by dividing 1000 milliseconds – 1 second – by 60). Therefore, the steps to increase and decrease the deadline, and the minimum deadline, were multiples of 17. That is, in case of a correct response, the deadline was decreased by 17 ms down to a minimum of 34 ms. In case of an error, the deadline was increased by 68 ms up to a maximum of 2000 ms). The maximum deadline of 2000 ms was selected based on the RT distributions of similar tasks in our previous work (Rey-Mermet et al., 2018). We set the starting value of the presentation time based on our pilot studies. To compensate for practice and/or fatigue effects, we repeated the calibration procedure before each experimental block. In each experimental block, the deadline was fixed at the median deadline of the previous calibration block. This median deadline was computed on all calibration trials, excluding warm-up trials (see Appendix C for the block structure and number of trials for each task).

Color Stroop task. Participants were asked to indicate the color of color words while ignoring the meaning of the words (MacLeod, 1991). To respond, they were instructed to press colored keys (i.e., keys *6*, *7*, *8*, and *9* of the keyboard which were covered with a red, blue, green, and yellow sticker, respectively). Participants were asked to use the index and middle fingers of the left and right hands. The stimuli were the German color words for red, blue, green, and yellow (i.e., “rot”, “blau”, “grün”, and “gelb”), and a sequence of five x characters (“xxxxx”). All stimuli were lowercase, displayed either in red, blue, green, or yellow. Trials were either incongruent, congruent, or neutral. In incongruent trials, the color did not correspond to the word meaning (e.g., the word “red” printed in blue). In congruent trials, the color corresponded to the word meaning (e.g., the word “red” printed in red). In neutral trials, the stimuli were colored “xxxxx” (i.e., stimuli without word meaning).

To familiarize with the task, participants first performed a practice block (16 trials) in which incongruent and congruent trials were presented randomly and with equal frequency. During each trial of the practice block (see Figure 1), a fixation cross was presented centrally for 500 ms. Then, the stimulus was presented centrally until the response, or until 2000 ms elapsed. After stimulus presentation, an accuracy feedback was presented in the middle of the screen for 500 ms, which was followed by a blank screen for 500 ms. During the entire trial sequence, stimulus-response mappings were presented in the lower part of the screen, and a visual timer was presented in the top part of the screen. The timer consisted of a white rectangle which gradually turned black from left to right and thus informed participants about the diminishing time left. Participants were informed that they had to respond before the rectangle turned entirely black. The accuracy feedback was a white happy smiley (2.54 cm x 2.54 cm) after a correct response, a white sad smiley (2.54 cm x 2. 54 cm) after an error, and the words “too slow” (in German: “zu langsam”) in case of no response within the deadline.

After this practice block, participants performed a calibration block (28 trials) in which only neutral trials were presented. During each trial of this block, a fixation cross was presented centrally for 500 ms. Then, the stimulus was presented centrally until the deadline, followed by a blank screen for 500 ms. We calibrated the deadline individually to achieve 75% correct responses with an initial deadline set at 650 ms. Participants were informed about the deadline of each trial with the visual timer. Neither stimulus-response mappings nor feedback were presented.

After this calibration block, participants performed an experimental block in which incongruent and congruent trials were presented randomly intermixed and with equal frequency. The trial procedure was similar to the calibration block, except that the deadline was fixed individually. Overall, participants repeated the sequence of calibration and experimental blocks for three times (see Table C1 in Appendix C for a description of the block structure of this task). Given the good reliability estimates for the tasks assumed to measure executive control in our previous study (Rey-Mermet et al., 2018), we decided to keep the same number of trials (i.e., 96 trials for each incongruent and congruent trials). Thus, the color Stroop task, comprised 192 trials presented in three experimental blocks. The dependent measure for this task was the difference in error rates (in %) between incongruent and congruent trials in the experimental blocks.

Number Stroop task. Participants were asked to determine the number of 1 to 4 centrally displayed characters while ignoring the numerical value of digit characters (Salthouse & Meinz, 1995). To respond, they were instructed to press the corresponding keys *1*, *2*, *3* or *4* on the upper row of the keyboard with the index and middle fingers of the left and right hands, respectively. In incongruent trials, the number of digits did not correspond to the digits displayed (e.g., 222). In congruent trials, the number of digits corresponded to the digits displayed (e.g., 22). In neutral trials, unrelated symbols were displayed (e.g., $$). The block sequence (see Table C1), the trial sequence (see Figure 1) and the dependent measure were the same as for the color Stroop task (except that based on our pilots, we set the initial deadline for the first calibration block to 600 ms).

Arrow flanker task. Participants were asked to respond to the direction of the central arrow (left or right) while ignoring four flanking characters (e.g., Unsworth & Spillers, 2010). To respond, they were instructed to press the keys *A* or *L* with the index fingers of the left or right hand, respectively. In incongruent trials, the central arrow indicated the opposite direction of the flanking arrows (e.g., <<><<). In congruent trials, the central arrow indicated the same direction of the flanking arrows (e.g., >>>>>). In neutral trials, unrelated symbols were displayed as flankers (e.g., ==>==). The block sequence (see Table C1), the trial sequence (see Figure 1) and the dependent measure were the same as for the color Stroop task (except that the initial deadline for the first calibration block was set at 400 ms).

Letter flanker task. Participants were asked to decide whether a centrally presented target was a vowel (E or U) or consonant (S or H) while ignoring four flanker characters (Eriksen & Eriksen, 1974). To respond, they were instructed to press the keys *A* or *L* with the index fingers of the left or right hand, respectively. In incongruent trials, the central letter belonged to the other category than the flanker letters (e.g., SSESS). In congruent trials, the central letter belonged to the same category as the flanker letters (e.g., UUEUU or UUUUU). In neutral trials, unrelated symbols (i.e., # or %) were displayed as flankers (e.g., ##U##). The block sequence (see Table C1), the trial sequence (see Figure 1) and the dependent measure were the same as for the color Stroop task (except that the initial deadline for the first calibration block was set at 600 ms).

Simon task. Participants were asked to decide whether a circle – presented on either the left or right side of the screen – was filled or unfilled while ignoring the position of the circle on the screen (Hommel, 2011). Participants were instructed to press the keys *A* or *L* with the index fingers of the left or right hand, respectively. In incongruent trials, the response position did not correspond to the position of the circle on the screen (e.g., a filled circle presented on the right side and requiring to press the left key *A*). In congruent trials, the left-right position of the response key corresponded to the position of the circle on the screen (e.g., a filled circle presented on the left side and requiring to press the left key *A*). The block sequence (see Table C1), the trial sequence (see Figure 1) and the dependent measure were the same as for the color Stroop task, except for the following modifications. First, the deadline for the first calibration block was set at 400 ms. Second, congruent trials – instead of neutral trials – were used in the calibration blocks. We opted for congruent trials in the calibration blocks because neutral trials are not typically used in the Simon task (Hommel, 2011).

Antisaccade task. In the antisaccade task (adapted from Friedman & Miyake, 2004; Rey-Mermet et al., 2018), participants were asked to indicate the direction of a small arrow (either left, up, or right). Arrows were presented briefly on the left or the right side of the screen and then masked. Shortly before the onset of the arrow, a black square appeared either to the right or to the left of the screen. In prosaccade trials, the square and the arrow appeared on the same side of the screen, whereas on antisaccade trials, the square and the arrow appeared on opposite sides of the screen. For antisaccade trials, participants were asked to inhibit a reflexive saccade toward the square and instead make a voluntary saccade to the opposite side in order to identify the briefly appearing target arrow. To ensure that participants performed saccades (and not head movements), a chin rest was used. This was approximately 57 cm away from the monitor.

During each trial (see Figure 1), a fixation cross appeared centrally for a variable amount of time (one of eight time intervals between 1500 and 3250 ms in 250-ms intervals, selected pseudo-randomly with equal probability and no repetition). A black square (0.32 cm x 0.32 cm) then appeared on one side of the screen for 166 ms, followed by the target stimulus, that is, an arrow inside of an open rectangle (1.6 cm x 1.6 cm). Both the square and the target were presented on either the left or right side of the screen (i.e., 9.5 cm away from the center). Size and eccentricity of the arrow were selected so that the arrow direction could not be identified while fixating the screen center. The target was followed by a mask (i.e., three arrows indicating left, up and right inside of an open rectangle of 4.80 cm x 3.75 cm), which was presented for 300 ms. Then, a blank screen was presented until a response was given or until the deadline (i.e., 1700 ms minus the target presentation time of the current trial). Participants responded by pressing the arrow key on the computer keypad that corresponded to the target with the index, middle and ring fingers of the right hand. In case of the practice block, the accuracy feedback appeared centrally for 500 ms, followed by a blank screen for 500 ms. In case of calibration and experimental blocks, only the blank screen was presented for 500 ms.

We calibrated the target presentation times (i.e., the time between onset of the arrow and onset of the mask) in prosaccade blocks so that participants achieved 80% correct responses. This cut-off was consistent with our previous study (Rey-Mermet et al., 2018) in which it worked well. Thus, in case of a correct response, the target presentation time was decreased by 17 ms (to a minimum of 34 ms). In case of an error, the target presentation time was increased by 85 ms (to a maximum of 740 ms). For all participants, the initial target presentation was set at 150 ms. After a calibration block with only prosaccade trials, participants completed an experimental block with only antisaccade trials in which target presentation time was fixed individually to the median target presentation times of the previous prosaccades blocks. Given the good reliability estimates for the antisaccade task in our previous study (Rey-Mermet et al., 2018), we decided to keep the same number of trials (i.e., 96 antisaccade trials). These were presented in two experimental blocks. The detailed description of block procedure is presented in Table C1. The dependent measure was the difference in error rates (in %) between antisaccade and prosaccade trials.

Stop-signal task. Participants were asked to classify pictures as representing living or non-living objects by pressing the keys *A* and *L* with the index fingers of the left and right hands, respectively. Twenty-four pictures were selected from Cycowicz, Friedman, Rothstein, and Snodgrass (1997). These pictures were presented with a grey frame, which on *stop* trials turned pink after a variable interval (i.e., the stop-signal delay, SSD) following the onset of the picture. Participants were instructed to withhold a response when the frame of the pictures turned pink (i.e., the stop-signal; see Logan, 1994).

To familiarize themselves with the pictures, participants first performed a practice block in which all pictures were presented once with a grey frame (i.e., only *go* trials were presented). During each trial of the practice block (see Figure 1), a fixation cross appeared centrally for 500 ms. Then, the target picture (9.22 cm x 9.05 cm) was presented centrally until a response was given or 2000 ms elapsed. After the target picture, the accuracy feedback was presented centrally for 500 ms, followed by a blank screen for 500 ms.

After the practice block, participants performed four experimental blocks in which *stop* trials were presented on 33% of the trials (see Table C1 in Appendix C for a detailed description of the block structure). The trial sequence was similar to the trial sequence for the practice block, except for the following two modifications (see Figure 1). First, no stimulus-response mappings were presented. Second, to prevent participants from waiting for the stop signal to occur, the feedback was introduced as a game in which the goal was to win as many points as possible. More precisely, on a *go* trial, if the response was fast enough (i.e., faster than the mean RT computed across the *go* trials from the previous block), participants were informed by one green happy smiley that they won one point. If the response was too slow (i.e., slower than the mean RT plus one standard deviation computed across the *go* trials from the previous block), participants were informed with one yellow neutral smiley that they lost one point. If the response was wrong, participants were instructed with one red sad smiley that they lost one point. On a *stop* trial, if no response was given, participants were informed with three green happy smileys that they won three points. In contrast, if a response was given, participants were informed with three red sad smileys that they lost three points.

In the stop-signal task, performance is modeled as a race between a *go* process, which is triggered by the presentation of the *go* stimulus, and a *stop* process, which is triggered by the presentation of the stop signal (Logan, 1994; Verbruggen & Logan, 2008). When the *stop* process finishes before the *go* process, the response is inhibited; when the *go* processes finishes before the *stop* process, the response is executed. Typically, executive control is measured by the latency of the *stop* process, referred to as the stop-signal RT (SSRT). In the present study, however, we aimed to measure performance on the stop-signal task through accuracy. Therefore, we adjusted the time between the onset of the picture and the stop-signal (i.e., the SSD) based on the RTs from the *go* trials of the previous block. That is, based on our pilot studies as well as on the data from Rey-Mermet et al. (2018), we computed the SSD with the following equation:

$$\text{SSD} = (2/3 * \text{median RT}_{go\ \text{trials of previous block}}) + \delta,$$

where δ is a variable time interval sampled pseudo-randomly with equal probability (excluding repetitions) between -200 and + 200 in intervals of one screen refresh cycle (17 ms). We found that this equation yielded SSD values for which the chance of stopping was comfortably away from floor and ceiling, thereby providing a sensitive measure of stopping ability. With that equation, the SSD was adjusted to the *go* RTs of each participant so that the participant could successfully stop their reaction in some *stop* trials but not in others. Thus, failure to inhibit the response on *stop* trials should result in higher error rates. Accordingly, the dependent measure was the error rates on *stop* trials.

gF tasks. In all tasks measuring gF, the dependent measure was error rates (in %). That is, the number of incorrectly or not solved items was divided by the total number of items. Block structure for each task is presented in Appendix C.

Letter sets test. Five sets of four letters were presented that all followed a certain logical pattern except for one set (Ekstrom, French, Harman, & Dermen, 1976). Participants had to select the deviating letter set. Participants had 7 minutes to complete each of the two test blocks.

Locations test. Participants had to discover the rule of patterns of dashes, one of which in each of four rows was replaced by an “x” (Ekstrom et al., 1976). Participants had to select the correct location of the next “x” out of five alternatives. Participants had 6 minutes to complete each of the two test blocks.

Raven's advanced progressive matrices (RAPM) test. Participants had to complete a pattern by choosing one of eight alternatives (Raven, 1990). We used the Arthur and Day’s short version (1994; see also Arthur, Tubre, Paul, & Sanchez-Ku, 1999). Participants had 15 minutes to complete the test.

Relationships test. Participants had to choose which out of five diagrams represents best a set of three given nouns (Ekstrom et al., 1976). For example, the set “animals, cats, and dogs” would be best represented by one circle corresponding to “animals” containing two separate circles for “cats” and “dogs”. Participants had 4 minutes to complete each of the two test blocks.

Syllogisms test. The task was to decide whether conclusions drawn from two premises with nonsensical content were logically valid (Ekstrom et al., 1976). For example, following the premises “all trees are fish” and “all fish are horses”, it would be logically correct to conclude that “therefore all trees are horses”. Participants had 4 minutes to complete each of the two test blocks.

WMC tasks. WMC was measured with four tasks, that is, two updating tasks and two complex span tasks. Each updating and complex span task was either numerical or spatial. In all tasks, the dependent measure was the error rates (in %). Block structure for each task is presented in Appendix C.

Numerical updating task. Four digits (ranging from 1 to 9) were presented in four different colors (i.e., red, blue, green, and orange; von Bastian et al., 2016). Participants had to memorize the digits in each color during 5000 ms, followed by a blank screen for 250 ms. In the subsequent updating step, a new digit was presented in one of the four colors. Participants were asked to update the value of the item with the corresponding color. Seven updating steps were presented for 1250 ms, followed by a blank screen for 250 ms. After these updating steps, participants were asked to recall the most recent digit of each color. To ensure that the initial set of memoranda had to be encoded, in 5 out of 25 trials recall was probed immediately after the initial encoding. Performance on these immediate probes were not included in the computation of the dependent measure.

Spatial updating task. Four to six colored dots were presented in a 4 x 4 matrix (De Simoni & von Bastian, 2017). Participants were asked to memorize the positions of colored dots and update their position during nine updating steps. For each updating step, we indicated the new position of one of the colored dots by presenting the to-be-updated dot in the center of the screen, together with an arrow pointing in the direction of the required mental shift of that dot (i.e., either left, right, up, or down). For example, if the blue dot was presented with an arrow pointing left, participants were asked to update the position of the blue dot by shifting it to the left by one matrix cell. During a trial, a fixation cross was presented for 2000 ms and was followed by the memoranda. Memoranda were simultaneously presented for 500 ms per colored dot (e.g., four memoranda were presented for 2000 ms). Each updating step lasted 500 ms.

Numerical complex span task. Participants had to memorize 3 to 5 two-digit numbers (see von Bastian et al., 2016). Presentation of the memoranda was interleaved by a distractor task in which participants were asked to judge the veracity of equations (e.g., “1 + 3 = 5”). Each digit was presented for 1000 ms. Each distractor task was presented for 3000 ms. At the end of each trial, memoranda had to be recalled in correct serial order, followed by a blank screen for 500 ms. For this task, error rates were computed as the proportion of items not recalled at the correct position (partial-credit unit score; see Conway et al., 2005).

Spatial complex span task. Participants had to remember the position of sequentially presented dots in a 5 x 5 matrix (see von Bastian & Eschen, 2016). The number of dots ranged from 4 to 6. In the distractor task, four dots arranged in an L-shape were presented concurrently, and participants were asked to judge whether the pattern emerging from these dots was vertical or horizontal. The trial sequence and dependent measure were the same as for the numerical complex span task.

Time estimation task. Participants were asked to estimate the duration of a constant time interval by marking its end through a key press. On each trial, a diamond (4 cm x 4 cm) was presented for the duration of the interval, and participants were asked to press the space key with the index finger of the right hand, making the key press coincide with the offset of the diamond as precisely as possible. At the beginning of each block, trials included the visual timer on the top of the screen (i.e., trials with timer), and participants were instructed to use the gradual progression of the visual timer to anticipate when the diamond would disappear. On further trials (i.e., trials without timer), the timer was removed and participants had to continue to press the space bar in sync with the offset of the diamond. During each trial (see Figure 1), a fixation cross was presented centrally for 500 ms. Then, the diamond was presented centrally for a fixed presentation time. Offset of each diamond was followed by a blank screen for a variable amount of time (i.e., 1 out of 11 time intervals between 800 and 1140 ms in 34-ms intervals, selected pseudo-randomly with equal probability) before the onset of the next diamond. Participants performed three blocks. Stimulus presentation time was constant in each block but varied across blocks. That is, the stimulus was presented for 500 ms in the first block, 750 ms in the second block and 1000 ms in the third block (see Table C1 in Appendix C for a description of the block structure of this task).[1]

The goal of this task was to assess individual differences in the ability to estimate the duration of the time intervals that were given as response intervals in the executive-control tasks. Individuals who are better in estimating these time intervals could make better use of the available time by delaying their response until just before the deadline without surpassing it. Therefore, time interval estimation ability could be confounded with our measures of executive control, and we aimed to control for that ability through the time estimation task. The measure of interest in this task is, thus, how precisely participants could anticipate the end of the time interval. Accordingly, the dependent measure for this task was the response time precision (i.e., the absolute difference between the stimulus offset time and the RT). This response time precision was computed separately for the trials with timer and for those without timer.

Procedure

Participants were tested in groups of up to five during two sessions lasting approximately 2-2.5 hours each (including breaks after every block and a longer break in the middle of each session). Both sessions were separated by at least 12 hours and maximally one week. In the first session, after informed consent was obtained, participants started with a questionnaire assessing demographics and then performed half of the tasks. In the second session, participants completed the Beck Depression Inventory II (Hautzinger, Keller, & Kühner, 2006) and then performed the remaining tasks.

In the first session, the tasks were ordered as follows: the letter sets test, the arrow flanker task, the numerical updating task, the stop-signal task, the spatial complex span task, the Simon task, and the RAPM test. In the second session, the tasks were ordered as follows: the relationships test, the color Stroop task, the spatial updating task, the locations test, the time estimation task, the numerical complex span task, the antisaccade task, the syllogisms test, and the letter flanker task. This task order was used for half of participants, and it was reversed for the other half to control for practice effects.

Data Preparation

For all constructs, mean error rates were computed as dependent measures for each participant and each task. To predict positive correlations between all measures, we coded the measures so that larger values indicate worse abilities.

For the tasks assumed to measure executive control (except the stop-signal task), we treated errors of commission (i.e., giving a wrong response) and errors of omission (i.e., failing to respond before the deadline) as equivalent. This decision follows necessarily from the rationale of measuring executive control through accuracy within a limited time window: We measure a person's ability to produce an accurate response within the allotted time. Any failure to do so, whether by giving a wrong response or by missing the deadline, reflects a lack of that ability. This way of scoring errors avoids contamination of the score with individual differences in speed-accuracy trade-off: People may differ in whether they prefer to respond fast at the price of committing more erroneous responses, or to respond slowly at the price of missing the deadline more often (i.e., committing more omission errors), but neither preference yields an undue advantage for the accuracy score. To remove any unsystematic variance involved in the calibration procedure, we computed a difference score between the mean error rates of antisaccade trials and those of prosaccade trials for the antisaccade task, and between the mean error rates of incongruent trials and those of congruent trials for the color and number Stroop, arrow and letter flanker, and Simon tasks. For the sake of completeness, error rates for incongruent and congruent trials are presented in Appendix D.

Model Estimation

Latent variable models were estimated in R using the lavaan package (Rosseel, 2012). Model fit was evaluated via multiple fit indices (Hu & Bentler, 1998, 1999): the χ^2 goodness-of-fit statistic, the Bentler's comparative fit index (CFI), the root mean square error of approximation (RMSEA), the standardized root-mean-square residual (SRMR), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). For the χ^2 statistic, a small, non-significant value indicates good fit. For the CFI, values larger than .95 indicate good fit, and values between .90 and .95 indicate acceptable fit. RMSEA values smaller than .06 and SRMR values smaller than .08 indicate good fit. However, as the RMSEA tends to over-reject true population models at small sample size (i.e., smaller than 250; Hu & Bentler, 1998), we report it for the sake of completeness only. For the AIC and the BIC, smaller indices indicate better fit.

To test if one model fit the data better than another, we performed two analyses. First, we conducted χ^2 difference ($\Delta\chi^2$) tests on nested models. If the more complex model (i.e., the model with more free parameters) yields a reduction in χ^2 that is significant given the loss of degrees of freedom, it is accepted as having better fit. Second, we performed a Bayesian hypothesis test using the BIC approximation (Wagenmakers, 2007). That is, we used the difference between the BIC for the null hypothesis (e.g., the single-factor model) and the BIC for the alternative hypothesis (e.g., the 2-factor model) to compute a Bayes factor (BF) in favor of the null hypothesis (BF_{01}) and in favor the alternative hypothesis (BF_{10}). Following Raftery's (1995) classification scheme, we considered a BF between 1-3 as weak evidence, between 3-20 as positive evidence, between 20-150 as strong evidence, and larger than 150 as very strong evidence. One advantage of using Bayesian hypothesis testing in addition to the more standard $\Delta\chi^2$ test was that we could assess the strength of evidence not only for the alternative but also for the null hypothesis.

In addition, the best-fitting models were considered as good only if the error variances were low and most factor loadings were significant and high (i.e., no factor should be dominated by the high loading of one task). Moreover, the amount of shared variance across the tasks had to be high. This "factor reliability" was measured with the Omega (ω) coefficient (Raykov, 2001).

Results

Results are reported in two steps. First, we examined the reliabilities and the correlational patterns of the scores derived from all tasks. Second, we used structural equation modeling (SEM) to measure each construct at the latent variable level and then to investigate the relations between the three constructs. We present the results from the calibration blocks in Appendix E. To determine whether participants with high WMC and gF abilities differed from participants with low WMC and gF abilities in how they handle the response deadline (for example by being better at meeting the response deadline and thus committing less errors of omission), we computed correlations between the error rates computed with both types of errors (i.e., errors of commission and omission) and the measures of WMC and gF. These results are presented in Appendix F.

Reliability and Correlations

As shown in Table 3, the reliability estimates of most task scores were good or at least acceptable, ranging from .58 (color Stroop) to .98 (spatial updating). Only the reliability of the letter flanker task was unacceptably low (.26). For this reason, we removed this measure from the subsequent SEM analyses.

The correlations are presented in Table 4. To assess the strength of each correlation, we also computed a Bayes Factor in favor of the alternative hypothesis (BF_{10}, i.e., in favor of a correlation) and a Bayes Factor in favor of the null hypothesis (BF_{01}, i.e., in favor of the absence of the correlation). These BFs are presented in Appendix G. Most correlations within and between the sets of measures assumed to assess gF and WMC were strong and significant. In line with these observations, most BF_{10} suggested positive to strong evidence for the correlations. In contrast, most correlations within the executive-control measures and between these measures and those assumed to assess gF and WMC were low and non-significant. This was further supported by most BF_{01}, suggesting positive to strong evidence for the absence of correlations. Furthermore, nearly all correlations between the executive-control measures and the time estimation task were low and non-significant, and most BF_{01} suggested positive to strong evidence for the absence of correlations. This speaks against the possibility that individuals who are better in estimating the time intervals would make better use of the available time in the executive-control tasks.

Structural Equation Modeling

Using SEM, we first replicated the relations between both gF and WMC as latent variables. Next, we intended to find a coherent factor of executive control and to investigate the relationships between executive control, gF and WMC.

WMC and gF as Distinct but Correlated Factors. To assess the relations between gF and WMC, we fit three models. Model 1 fit the two-factor model in which gF and WMC represent two distinct yet correlated factors, as depicted in Figure 2a. As shown in Table 5, this model provided an acceptable fit to the data. All tasks loaded significantly on their hypothesized factor, and error variances were relatively low (see Table 6). Factor reliability was high for both factors (i.e., $\omega = .71$ for gF, and $\omega = .76$ for WMC). The correlation between the gF and WMC factors was high (.77) and significant ($p < .001$, with 95 % confidence intervals (CIs) = [.65, .88]).

To ensure that the correlation between the factors was required, we fit Model 2 with the constraint that the factors were assumed to be orthogonal. As shown in Table 5, this model provided a poor fit to the data. The difference in χ^2 between Models 1 and 2 was significant, $\chi^2_{diff}(1) = 72.46$, $p < .001$, and the BF_{10} in favor of the alternative hypothesis (i.e., Model 1 ≠ Model 2) was very strong ($BF_{10} = 4.04e+14$). This indicates a better fit for Model 1 than for Model 2.

Model 3 fit a unity model, in which all tasks loaded on a single factor. This model also provided a poor fit to the data (see Table 5). Again, the difference in χ^2 between Models 1 and 3 was significant, $\chi^2_{diff}(1) = 25.62$, $p < .001$, and the BF_{10} in favor of the alternative hypothesis (i.e., Model 1 ≠ Model 3) was very strong ($BF_{10} = 2.72e+04$), indicating a better fit for Model 1 than for Model 3. Taken together, the results replicate previous findings (e.g., Engle & Kane, 2004; Krumm et al., 2009; Unsworth et al., 2014; Unsworth, Spillers, et al., 2009) by showing that gF and WMC represent correlated but distinct factors.

Executive Control and its Relations to gF and WMC. In the next step, we aimed to find a coherent factor of executive control. To this end, we fitted a Model 4 in which the tasks assumed to assess executive control loaded on a single factor. This model is depicted in Figure 2b. However, it did not converge.

To identify a coherent factor of executive control, we also fitted the following unitary models: (1) a Model 5 in which the color Stroop task was removed because this task had the lowest reliability among the remaining executive-control measures (see Table 3); (2) a Model 6 in which only the tasks used by Friedman an Miyake (2004) were included (i.e., the antisaccade, stop-signal, color Stroop, and arrow Flanker tasks); and (3) a saturated Model 7 in which only the tasks used by Miyake et al. (2000) to assess inhibition were included (i.e., the antisaccade, stop-signal, and color Stroop tasks). Model 5 did not converge. Models 6 and 7 had acceptable fit statistics (see Table 5). However, for both models, none of the factor loadings were significant, and all error variances were high (see Table 6). Factor reliability was very low for both models (i.e., $\omega = .02$ for Model 6 and $\omega = .07$ for Model 7). Therefore, the data are consistent with the assumption of a common factor, but that common factor explained only a very small fraction of the variance in each task. Hence, we could not establish a coherent factor of executive control. Given that outcome, we could not investigate the relations between all three constructs at the latent variable level.

To investigate the relationships between the executive-control measures and the constructs of gF and WMC, we therefore opted for a different strategy. For each individual measure of executive control, we examined its relationship to the gF and WMC factors. To this end, we fitted a Model 8 in which each executive-control measure predicted the gF and WMC factors (see Figure 2c). The goodness-of-fit statistics are summarized in Table 7. For all measures of executive control, the model provided an acceptable fit to the data. However, none of the regression coefficients between the individual measure of executive control and the gF or WMC factors, respectively, was significant. Moreover, we assessed the strength of each regression coefficient by comparing Model 8 to a Model 9 in which either the regression to gF or to WMC, respectively, was set to 0. Using the Bayesian hypothesis testing with the BIC approximation, we found that the BF_{01} in favor of the null hypothesis (i.e., Model 8 = Model 9) revealed positive evidence against a correlation between each executive-control task and the latent variables of gF and WMC (see Table 7). Taken together, these results challenge the hypothesis that executive control – even on the level of individual tasks – is related to gF and WMC.

To test for the robustness of these results, we re-ran the analyses by applying the following modifications: (1) instead of using the error rates difference between prosaccade and antisaccade trials for the antisaccade task, we followed previous studies by using as dependent measure the error rates on antisaccade trials; (2) instead of using the error rates on stop trials for the stop-signal task, we used the SSRT computed with the integration method[2]; (3) for the color Stroop, number Stroop, arrow flanker, and Simon tasks, incongruent trials were regressed on congruent trials; for the antisaccade task, antisaccade trials were regressed on prosaccade trials; and for all these measures, residuals were used as measures of executive control; (4) we estimated bi-factor models in which performance on antisaccade and prosaccade trials of the antisaccade task, and performance on incongruent and congruent trials of the color Stroop, number Stroop, arrow flanker, and Simon tasks, were forced to load on a baseline factor, and performance on antisaccade trials and on incongruent trials, as well as performance in the stop-signal task, were forced to load on an executive-control factor (see Figure 2d for an illustration); (5) we removed the impact of practice and/or fatigue effects by computing a linear regression analysis with participant's mean error rates as outcome variable and task-order (forward, backward) as a dichotomous predictor (coded as -0.5 and 0.5, respectively), and we used the residuals as dependent measures; (6) we included all participants (including those with missing data) and we ran the structural equation modeling with case-wise maximum likelihood; (7) we checked for multivariate normality in the *experimental* blocks using Mardia's (1970) kurtosis index and we removed multivariate outliers (i.e., cases with significant Mahalanobis's d2 values); (8) we checked for multivariate normality in the *calibration* blocks using Mardia's (1970) kurtosis index and we removed multivariate outliers (i.e., cases with significant Mahalanobis's d2 values) when performing the analyses on the experimental blocks; (9) we removed 40 participants because of a score above 13 in the BDI-II; and (10) we removed one participant who had error rates close to the chance level across several tasks and who postponed her/his reactions in several calibration blocks. The goodness-of-fit statistics of these model assessments and the parameter estimates can be found at https://osf.io/4t4h5. None of the competitive fitting of SEM models resulted in a best-fitting model with a high amount of shared variance among the executive-control tasks (i.e., with a high factor reliability, low error variances, and significant but non-dominant factor loadings).

General Discussion

The present study had two purposes. The first goal was to find a coherent latent variable of executive control. To maximize the chance of finding such a factor, we measured executive control through accuracy under a time limit. We decided to do so because this dependent variable reflects a person's ability to do a task quickly and accurately in a single dependent variable, therefore not being compromised by individual differences in speed-accuracy trade-offs. In addition, by calibrating the available time according to each person's performance in the baseline condition, we obtained measures of executive control that are not confounded with individual differences in baseline performance in the way that differences between RTs are. The second goal was to investigate the relationships between executive control, gF, and WMC. Measuring all three constructs – including executive control – on the same scale (i.e., accuracy) should increase the chance of finding substantial relations between all these constructs because the difference between constructs would not be confounded with potential differences in method variance.

Although executive control and inhibition are not isomorphic constructs, we measured executive control with tasks assumed to assess inhibition (i.e., the Stroop, flanker, Simon, antisaccade, and stop-signal tasks) for two reasons. First, these tasks are most frequently used to assess executive control (e.g., Chuderski et al., 2012; Kane et al., 2016; Redick et al., 2016; Shipstead et al., 2014; Unsworth & Spillers, 2010). Second, inhibition tasks have been assumed to provide the purest measure of the basic ability involved in all forms of executive control (Friedman & Miyake, 2016; Miyake & Friedman, 2012; Munakata et al., 2011). Therefore, if the core construct of executive function falls apart, as our results suggest, then this questions the unity of the executive-function construct as a whole.

The results showed good reliabilities of our measures of executive control – in most cases as good as, or even better than, those found in previous research (e.g., Friedman & Miyake, 2004; Krumm et al., 2009; Pettigrew & Martin, 2014; Rey-Mermet et al., 2018; Shipstead et al., 2014; Stahl et al., 2014; Unsworth et al., 2014; Unsworth, Spillers, et al., 2009; von Bastian et al., 2016). However, these measures correlated neither amongst each other nor with the gF and WMC tasks. Bayesian inference revealed positive to substantial evidence for the absence of these correlations. SEM identified no model including executive control as a coherent latent variable. Measures of WMC and gF could be best accommodated by a model with two distinct but correlated factors for WMC and gF. However, these latent constructs were not related to the individual measures of executive control.

Measuring Executive Control through the Response Deadline Procedure

In the present study, we measured executive control trough accuracy by implementing a limited response time window. The goal in using this method was to map individual differences in both processing speed and processing accuracy onto a single measurement scale. This is a new method of measurement, and as such there is the risk that some of its novel features undermined the measures' validity. We discuss three risk factors potentially limiting our measures' validity before addressing the question of validity more generally.

First, in most of our tasks, the response time window was presented with a timer on the top of the screen. One may argue that presenting a visual timer encourages participants to divide their visual attention between the critical stimuli and the timer, so that performance reflects in part people's ability to divide attention. There are three reasons speaking against this possibility. (1) If participants divided their attention between the main task and the timer, this would be a common process shared among all our executive-control measures. If anything, this should have increased the chance of finding systematic positive correlations between them, which, however, were not observed (see Table 4 and Appendix G). (2) If participants divided their attention, we should have found a correlation between the executive-control measures and the time-estimation task in which the timer was presented. In this task, participants were instructed to pay attention to both the stimulus and the timer, and thus they had to divide their attention. However, the results showed no such correlations for most executive-control measures (see Table 4 and Appendix G). (3) The stop-signal task and the antisaccade task did not include a timer. Thus, if presenting the timer somehow undermined the validity of the executive-control measures, we would have found correlations between these executive-control tasks and the tasks measuring WMC and gF. However, we found no correlations even for these tasks (see Table 4 and Appendix G).

A second limitation concerns the small number of calibration trials used in each executive-control task. Although most participants' accuracies in the calibration blocks were reasonably close to the criterion, there was still substantial variance left (see Figure E1). We interpreted this residual variance as measurement noise, and we computed a difference measure between incongruent and congruent trials to remove it. This interpretation was based on the assumption that successful calibration should remove systematic, true variance between individuals, and on the fact that the reliability of accuracy performance in the calibration blocks was low (see Table E1). However, the low reliability may be the result of the small number of calibration trials. Therefore, future work might improve the validity of the deadline-based measures by using a longer calibration period that more thoroughly removes all systematic sources of variance in the baseline conditions.

A third limitation is that we used also only a small number of practice trials for each task. As a consequence, participants might not have learned the stimulus-response mappings sufficiently well before testing commenced. Learning the mappings might have continued to a substantial degree during the test phase, and thereby individual differences in learning efficiency could have contaminated the executive-control measures. This might be particularly true for tasks in which the stimulus-response mapping was arbitrary, such as the color Stroop, letter flanker, and Simon tasks. Future studies could investigate whether increasing the number of practice trials results in executive-control measures that correlate more systematically with each other. It should be noted, however, that against this possibility, some authors have argued that task novelty is a requirement for a valid executive-function measure (Chan, Shum, Toulopoulou, & Chen, 2008; Rabbitt, 1997)

Despite our efforts to minimize all known threats to the validity of the executive-function measures we used – in particular, individual differences in speed-accuracy trade-off, differences in processing speed, and differences in episodic-memory contributions to interference control – we acknowledge that we cannot offer positive evidence for the validity of our measures. Where could positive evidence come from? A first source of construct validation would be to show positive correlations between different measures of the hypothetical construct. We were unable to find such positive correlations in a consistent manner. This negative result is ambiguous because it could mean that the hypothetical construct does not exist, but it could also mean that the indicators used to measure it lack validity. A second source of construct validation would be to demonstrate positive correlations between our new measures and already established measures of executive control. For instance, we could ask whether the measures derived from our deadline method correlate with conventional RT-difference measures from the same experimental paradigms. This approach is premised on the validity of established executive-control measures. However, we developed our novel approach precisely because the validity of conventional RT-based measures has been questioned. Therefore, they cannot serve as solid criteria for validation. If we found poor correlations between our response-deadline based measures and conventional RT-based measures of executive control, it could indicate poor validity of the former or the latter (or both). To conclude, the lack of positive evidence for the validity of our executive-control measures remains a limitation of the present work. It is a reflection of a principled problem of psychometrics: There is no way to establish the validity of a measurement instrument for a hypothetical construct separately from establishing the validity of the construct in question. As we discuss next, we suspect that the validity of the executive-control construct itself is lacking.

Does Executive Control Exist as a Psychometric Construct?

We were unable to establish a coherent factor reflecting executive control despite our efforts to measure executive control with improved methods that overcome potential drawbacks of previous studies. Some previous studies, however, have been successful in establishing executive control as a coherent factor (see Table 1). What explains this discrepancy?

One potential explanation is that our calibration method removed shared variance in processing speed from the executive-control measures. As explained in the introduction, due to the proportional nature of individual differences in general processing speed, using RT differences to measure executive control entails the risk of contaminating these measures with variance in general processing speed, which would artificially inflate the correlations between them. The calibration of response deadlines in the present study reduces that risk. We acknowledge, however, that we can offer only indirect evidence for the conjecture that executive-control measures in previous studies were contaminated with processing-speed variance. This assumption is based on the argument that general processing speed is expressed as a proportional slowing factor for RTs across experimental conditions (Zheng et al., 2000), so that slower individuals also show larger RT differences between any two conditions. One way to remove individual differences in proportional slowing factors is to use differences between log-transformed RTs as measures of executive control. When such measures were used, no coherent factor of inhibition could be established (Rey-Mermet et al., 2018). Additional empirical support for our conjecture comes from Jewsbury et al. (2016), who re-analyzed seven datasets (Brydges et al., 2012; Friedman & Miyake, 2004; Friedman et al., 2006, 2008; Hedden & Yoon, 2006; Miyake et al., 2000; van der Sluis et al., 2007), and showed that executive control – in particular, inhibition – does not separate from processing speed as conceptualized within the Cattell-Horn-Carroll (CHC) model of cognitive abilities.

Another potential reason for the discrepancy between our results and many previous studies might be that in the present study, as well as in Rey-Mermet et al. (2018), we took extra care to control for the impact of episodic memory and associative learning by excluding immediate stimulus repetitions and by avoiding stimulus overlap across tasks (e.g., Hommel, 2004; Mayr et al., 2003). Therefore, it is possible that reducing the contribution of these memory processes decreased the shared variance between the executive-control measures, which, in turn, reduced the correlations between them. In contrast, in most previous studies reported in Table 1, these memory processes might have been common to many or all executive-control measures, hence increasing the cross-task correlations. This may have happened even when difference scores were used as measures of executive control because these memory processes may contribute more to performance when responding to incongruent trials than when responding to congruent trials (see Verguts & Notebaert, 2008, 2009). In sum, it is possible that, in past studies, memory-related variance contributed to the difference scores used, and memory-related variance, not executive control, explained the covariance between the tasks. We acknowledge that this possibility is purely speculative so far. Future research could test it by investigating whether independent measures of episodic memory or associative learning explain shared variance among executive-control tasks if (and only if) the contributions of these processes to executive-control measures are not controlled.

In general, there are at least two explanations for the absence of an executive control factor. First, the tasks used to measure executive control are not appropriate to assess this ability. Thus, even though these tasks are well-established instruments to study executive control in experimental psychology, this might not be sufficient to produce good measures of individual differences in executive control. One reason for this paradox might be that these tasks had low performance variability between participants, which is optimal to find replicable effects in experimental psychology, but is clearly suboptimal in individual differences research (see

Cooper, Gonthier, Barch, & Braver, 2017; Hedge, Powell, & Sumner, 2017). The present results speak against this possibility because our measures of executive control – with one exception – had acceptable reliability, reflecting a substantial proportion of true variance between individuals. There remains the possibility discussed above that our measures, though reliable, lacked validity, and there is no way we can entirely rule this out.

The second possibility is that executive control – in particular, inhibition – might not exist as a psychometric construct (see Rey-Mermet et al., 2018; see also Karr et al., 2018). This means that although individuals differ reliably in their ability to apply executive control in specific tasks, these individual differences do not reflect differences in a more general executive control or inhibition ability. Thus, whatever the commonly used laboratory measures designed for studying executive control actually assess is highly task-specific. This interpretation would call into question the seminal executive-control model proposed by Miyake et al. (2000) and its updates (Friedman & Miyake, 2016; Friedman et al., 2008; Miyake & Friedman, 2012), as well as the experimental research of executive control, as it questions the validity of many of the measures used in that research, such as the Stroop, flanker, or Simon tasks. It would also provide no comfort for research on individual and developmental differences in executive control, which, in most part, use the same experimental tasks (e.g., Brydges et al., 2012; Hedden & Yoon, 2006; Hull et al., 2008; Pettigrew & Martin, 2014; van der Sluis et al., 2007). Moreover, this interpretation would limit the generality of computational neural network models of executive functions (Herd et al., 2014; Wiecki & Frank, 2013; see Munakata et al., 2011, for an overview).[3] These models provide impressive explanations of performance in executive-control tasks, but these explanations are, as yet, task-specific: It has not been demonstrated that they can explain behavior across multiple experimental paradigms with the same set of parameters.

Link of Executive Control to gF and WMC?

As the results revealed no latent variable of executive control, it was not possible to investigate the relationships between all three constructs at the latent variable level. Therefore, we opted for a different strategy by examining the relations between each individual measure of executive control and the latent constructs of gF and WMC. None of the executive-control measures was consistently found to be related to gF and WMC. Thus, even in a design in which the chance of finding such a relation was maximized because all constructs were measured on the same scale (i.e., accuracy), the findings did not support the hypothesis that executive control contributes to WMC or gF (e.g., Engle & Kane, 2004; Kane et al., 2001, 2007; Unsworth et al., 2014; Unsworth, Spillers, et al., 2009).

Most previous research, however, has shown an association between executive control and WMC or gF even when only one task was used to measure each construct of interest (see, e.g., Ahmed & Fockert, 2012; Heitz & Engle, 2007; Hutchison, 2011; Kane et al., 2001; Kane & Engle, 2003; Kiefer, Ahlegian, & Spitzer, 2005; Long & Prat, 2002; Mccabe, Robertson, & Smith, 2005; McVay & Kane, 2009; Meier & Kane, 2013, 2015; Morey et al., 2012; Poole & Kane, 2009; Redick, Calvo, Gay, & Engle, 2011; Redick & Engle, 2006; Unsworth, Redick, Spillers, & Brewer, 2012; Unsworth, Schrock, & Engle, 2004). Therefore, one may ask why our results differ from the findings of these studies. As discussed above, the reason might be that we measured executive control with improved methods that reduced the contamination of executive-control measures with variance due to general processing speed, and episodic memory. Our goal was to remove any other source of variance shared between tasks other than the one related to the efficiency of executive control. These efforts apparently have not only removed all sources of variance shared between different measures of executive control but also all sources of variance they shared with WM and gF.

One potential reason for concern about our study is the failure to replicate the correlation between the antisaccade task and WMC that has been obtained fairly consistently in previous studies (see Table 1). Different from most previous studies, here we controlled for the impact of non-executive-control processes by using a calibration procedure on the prosaccade block before measuring performance in the antisaccade block. In contrast, in previous research (e.g., Chuderski, 2014, 2015b; Kane et al., 2016; McVay & Kane, 2012; Meier, Smeekens, Silvia, Kwapil, & Kane, 2018; Shipstead et al., 2014; Unsworth et al., 2014; Unsworth & Spillers, 2010; Unsworth, Spillers, et al., 2009), participants were only tested on the antisaccade trials with a constant time window for detecting the target stimulus. In some studies, participants were asked to perform a few prosaccade trials (about 10-12) before performing the antisaccade trials (Redick et al., 2016; Unsworth & McMillan, 2014; Unsworth, Redick, Lakey, & Young, 2010; Unsworth & Spillers, 2010; Unsworth, Spillers, et al., 2009). Nevertheless, the dependent measure was simply the error rates on antisaccade trials (see Chuderski et al., 2012; Klauer et al., 2010, for exceptions). This dependent variable potentially confounds individual differences in the inhibition of a prosaccade with individual differences in the general speed of processing an only briefly presented stimulus. In intelligence research, processing speed has often been measured through the "inspection time" paradigm in which participants must make a perceptual judgment within a limited time window. Performance in this task has been found to correlate about r = .30 with gF (Kranzler & Jensen, 1989). The antisaccade task combines a task very similar to the inspection-time paradigm with the requirement to countermand a reflexive prosaccade. As such, accuracy in the antisaccade task could reflect a combination of variance in inhibition with variance in general processing speed, and it could be the latter that drives the correlation with gF and WMC.[4]

Finally, it should be noted that one study has found that presenting prosaccade trials before antisaccade trials reduced the relation with WMC such that individuals with low and high WMC no longer differed in antisaccade RT performance (Kane et al., 2001). Therefore, presenting prosaccade trials before antisaccade trials, as we did, might change the nature of the processes underlying antisaccade performance in ways that are as yet poorly understood. Taken together, although the antisaccade task seems a promising paradigm to measure executive control, further experimental work is called for to understand better what it measures, and care needs to be taken to prevent contamination of the dependent variable with general processing speed.

Conclusion

In the present study, we measured executive control, gF, and WMC through accuracy to create optimal conditions to find a coherent latent variable of executive control and substantial correlations between all three constructs. Despite satisfactory reliabilities, the measures of executive control hardly correlated among each other or with the measures of gF and WMC. No factor of executive control could be identified in SEM. Therefore, measuring executive control through accuracy does not overcome the difficulties with establishing an executive-control construct as a latent variable. Moreover, the individual measures of executive control were not consistently correlated with WMC and gF. These findings challenge (1) the existence of executive control as a psychometric construct and (2) the assumption that WMC and gF are closely related to the ability to supervise and control ongoing thoughts and actions.

References

Ahmed, L., & Fockert, J. W. de. (2012). Focusing on Attention: The Effects of Working Memory Capacity and Load on Selective Attention. *PLOS ONE*, *7*(8), e43101. https://doi.org/10.1371/journal.pone.0043101

Arthur, W., & Day, D. V. (1994). Development of a Short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement*, *54*(2), 394–403. https://doi.org/10.1177/0013164494054002013

Arthur, W., Tubre, T. C., Paul, D. S., & Sanchez-Ku, M. L. (1999). College-Sample Psychometric and Normative Data on a Short Form of the Raven Advanced Progressive Matrices Test. *Journal of Psychoeducational Assessment*, *17*(4), 354–361. https://doi.org/10.1177/073428299901700405

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*(1), 5–13.

Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence*, *40*(5), 458–469. https://doi.org/10.1016/j.intell.2012.05.006

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Theory and methodology of frontal and executive function* (pp. 81–116). East

Sussex, UK: Psychology Press. Retrieved from http://www.psypress.com/9780863774850

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://doi.org/10.1037/h0046743

Chan, R., Shum, D., Toulopoulou, T., & Chen, E. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201–216. https://doi.org/10.1016/j.acn.2007.08.010

Chuderski, A. (2014). The relational integration task explains fluid reasoning above and beyond other working memory tasks. *Memory & Cognition*, *42*(3), 448–463. https://doi.org/10.3758/s13421-013-0366-x

Chuderski, A. (2015a). High intelligence prevents the negative impact of anxiety on working memory. *Cognition and Emotion*, *29*(7), 1197–1209. https://doi.org/10.1080/02699931.2014.969683

Chuderski, A. (2015b). The broad factor of working memory is virtually isomorphic to fluid intelligence tested under time pressure. *Personality and Individual Differences*, *85*, 98–104. https://doi.org/10.1016/j.paid.2015.04.046

Chuderski, A., & Necka, E. (2012). The contribution of working memory to fluid reasoning: Capacity, control, or both? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1689–1710. https://doi.org/10.1037/a0028465

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*(3), 278–295. https://doi.org/10.1016/j.intell.2012.02.010

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

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

Cooper, S. R., Gonthier, C., Barch, D. M., & Braver, T. S. (2017). The Role of Psychometrics in Individual Differences Research in Cognition: A Case Study of the AX-CPT. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01482

Cowan, N. (1998). Visual and auditory working memory capacity. *Trends in Cognitive Sciences*, *2*(3), 77. https://doi.org/10.1016/S1364-6613(98)01144-9

Crawford, T. J., Parker, E., Solis-Trapala, I., & Mayes, J. (2011). Is the relationship of prosaccade reaction times and antisaccade errors mediated by working memory? *Experimental Brain Research*, *208*(3), 385–397. https://doi.org/10.1007/s00221-010-2488-8

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*(2), 171–237.

Dang, C.-P., Braeken, J., Colom, R., Ferrer, E., & Liu, C. (2014). Why is working memory related to intelligence? Different contributions from storage and processing. *Memory*, *22*(4), 426–441. https://doi.org/10.1080/09658211.2013.797471

De Simoni, C., & von Bastian, C. C. (2017). No evidence for effects of updating and binding training on working memory capacity and efficiency. *Manuscript Submitted for Publication*.

Dennis, I., & Evans, J. S. B. T. (1996). The speed-error trade-off problem in psychometric testing. *British Journal of Psychology*, *87*(1), 105.

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 170–189.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

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

Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. In B.-P. of L. and Motivation (Ed.) (Vol. 44, pp. 145–199). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/S007974210344005X

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target 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 interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., & Miyake, A. (2016). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*. https://doi.org/10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Gignac, G. E., & Kretzschmar, A. (2016). Endorsing a correlated-factor model: Not as simple as it seems. *Manuscript Submitted for Publication*.

Guye, S., & von Bastian, C. C. (2017). Working memory training in older adults: Bayesian evidence supporting the absence of transfer. *Psychology and Aging*, *32*(8), 732–746. https://doi.org/10.1037/pag0000206

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Beck Depression Inventar II (BDI 2)*. Frankfurt: Harcourt Test Service.

Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology*, *20*(5), 511–528. https://doi.org/10.1037/0894-4105.20.5.511

Hedge, C., Powell, G., & Sumner, P. (2017). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 1–21. https://doi.org/10.3758/s13428-017-0935-1

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, *136*(2), 217–240. https://doi.org/10.1037/0096-3445.136.2.217

Herd, S. A., O′Reilly, R. C., Hazy, T. E., Chatham, C. H., Brant, A. M., & Friedman, N. P. (2014). A neural network model of individual differences in task switching abilities. *Neuropsychologia*, *62*, 375–389. https://doi.org/10.1016/j.neuropsychologia.2014.04.014

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11), 494–500. https://doi.org/10.1016/j.tics.2004.08.007

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica*, *136*(2), 189–202. https://doi.org/10.1016/j.actpsy.2010.04.011

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior Research Methods*, *46*(3), 702–721. https://doi.org/10.3758/s13428-013-0411-5

Hull, R., Martin, R. C., Beier, M. E., Lane, D., & Hamilton, A. C. (2008). Executive function in older adults: A structural equation modeling approach. *Neuropsychology*, *22*(4), 508–522. https://doi.org/10.1037/0894-4105.22.4.508

Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 851–860. https://doi.org/10.1037/a0023437

Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell—Horn—Carroll model. *Journal of Experimental Psychology: General*, *145*(2), 220–245. https://doi.org/10.1037/xge0000119

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. A., & 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., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. *Variation in Working Memory*, *1*, 21–48.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. https://doi.org/10.1037/0096-3445.132.1.47

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

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General*, *145*(8), 1017–1048. https://doi.org/10.1037/xge0000184

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*. https://doi.org/10.1037/bul0000160

Keye, D., Wilhelm, O., Oberauer, K., & Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: testing means and covariance hypothesis about the Simon and the Eriksen Flanker task. *Psychological Research*, *73*(6), 762–776. https://doi.org/10.1007/s00426-008-0188-9

Keye, D., Wilhelm, O., Oberauer, K., & Ravenzwaaij, D. (2010). Individual differences in conflict-monitoring: testing means and covariance hypothesis about the Simon and the Eriksen Flanker task - Erratum. *Psychological Research*, *74*(2), 237–238. https://doi.org/10.1007/s00426-009-0257-8

Keye, D., Wilhelm, O., Oberauer, K., & Stürmer, B. (2013). Individual differences in response conflict adaptations. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00947

Kiefer, M., Ahlegian, M., & Spitzer, M. (2005). Working Memory Capacity, Indirect Semantic Priming, and Stroop Interference: Pattern of Interindividual Prefrontal Performance Differences in Healthy Volunteers. *Neuropsychology*, *19*(3), 332–344. https://doi.org/10.1037/0894-4105.19.3.332

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *The Quarterly Journal of Experimental Psychology*, *63*(3), 595–619. https://doi.org/10.1080/17470210903076826

Kranzler, J. H., & Jensen, A. R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence*, *13*(4), 329–347. https://doi.org/10.1016/S0160-2896(89)80006-6

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence*, *37*(4), 347–364. https://doi.org/10.1016/j.intell.2009.02.003

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). San Diego, CA, US: Academic Press.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*(2), 393–434. https://doi.org/10.1037/0033-295X.108.2.393

Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: An individual differences investigation. *Memory & Cognition*, *30*(2), 294–301. https://doi.org/10.3758/BF03195290

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519–530. https://doi.org/10.1093/biomet/57.3.519

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, *6*(5), 450–452. https://doi.org/10.1038/nn1051

Mccabe, D. P., Robertson, C. L., & Smith, A. D. (2005). Age Differences in Stroop Interference in Working Memory. *Journal of Clinical and Experimental Neuropsychology*, *27*(5), 633–644. https://doi.org/10.1080/13803390490919218

McCabe, D. P., Roediger, H. L. I., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222–243. https://doi.org/10.1037/a0017619

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 196–204. https://doi.org/10.1037/a0014104

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, *141*(2), 302–320. https://doi.org/10.1037/a0025250

Meier, M. E., & Kane, M. J. (2013). Working memory capacity and Stroop interference: Global versus local indices of executive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 748–759. https://doi.org/10.1037/a0029200

Meier, M. E., & Kane, M. J. (2015). Carving executive control at its joints: Working memory capacity predicts stimulus-stimulus, but not stimulus-response, conflict. *Journal Of Experimental Psychology. Learning, Memory, And Cognition*, *41*(6), 1849–1872. https://doi.org/10.1037/xlm0000147

Meier, M. E., Smeekens, B. A., Silvia, P. J., Kwapil, T. R., & Kane, M. J. (2018). Working memory capacity and the antisaccade task: A microanalytic–macroanalytic investigation of individual differences in goal activation and maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 68–84. https://doi.org/10.1037/xlm0000431

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Miyake, A., & Shah, P. (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press.

Morey, C. C., Elliott, E. M., Wiggers, J., Eaves, S. D., Shelton, J. T., & Mall, J. T. (2012). Goal-neglect links Stroop interference with working memory capacity. *Acta Psychologica*, *141*(2), 250–260. https://doi.org/10.1016/j.actpsy.2012.05.013

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, *15*(10), 453–459. https://doi.org/10.1016/j.tics.2011.07.011

Oberauer, K. (2009). Design for a Working Memory. In B.-P. of L. and Motivation (Ed.) (Vol. 51, pp. 45–100). Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/S007974210951002X

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626. https://doi.org/10.1016/j.jml.2006.08.009

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity — facets of a cognitive ability construct. *Personality and Individual Differences*, *29*(6), 1017–1045. https://doi.org/10.1016/S0191-8869(99)00251-2

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Pettigrew, C., & Martin, R. C. (2014). Cognitive declines in healthy aging: Evidence from multiple aspects of interference resolution. *Psychology and Aging*, *29*(2), 187–204. https://doi.org/10.1037/a0036085

Poole, B. J., & Kane, M. J. (2009). Working-memory capacity predicts the executive control of visual search among distractors: The influences of sustained and selective attention.

Quarterly Journal of Experimental Psychology, *62*(7), 1430–1454. https://doi.org/10.1080/17470210802479329

Rabbitt, P. (1997). *Methodology of frontal and executive function*. East Sussex, UK: Psychology Press.

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063

Raven, J. C. (1990). *Advanced Progressive Matrices: Sets I, II*. Oxford, U.K: Oxford Psychologists Press.

Raykov, T. (2001). Estimation of congeneric scale reliability using covariance structure analysis with nonlinear constraints. *British Journal of Mathematical and Statistical Psychology*, *54*(2), 315–323. https://doi.org/10.1348/000711001159582

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: Selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 308–324. https://doi.org/10.1037/a0022216

Redick, T. S., & Engle, R. W. (2006). Working memory capacity and attention network test performance. *Applied Cognitive Psychology*, *20*(5), 713–721. https://doi.org/10.1002/acp.1224

Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., … Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General*, *145*(11), 1473–1492. https://doi.org/10.1037/xge0000219

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

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *50*(6), 297–306.

Schachar, R., Logan, G. D., Robaey, P., Chen, S., Ickowicz, A., & Barr, C. (2007). Restraint and cancellation: Multiple inhibition deficits in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, *35*(2), 229–238. https://doi.org/10.1007/s10802-006-9075-2

Schaeffer, D. J., Chi, L., Krafft, C. E., Li, Q., Schwarz, N. F., & McDowell, J. E. (2015). Individual differences in working memory moderate the relationship between prosaccade latency and antisaccade error rate: Cognitive control and working memory. *Psychophysiology*, *52*(4), 605–608. https://doi.org/10.1111/psyp.12380

Schweizer, K., & Moosbrugger, H. (2004). Attention and working memory as predictors of intelligence. *Intelligence*, *32*(4), 329–347. https://doi.org/10.1016/j.intell.2004.06.006

Schweizer, K., Moosbrugger, H., & Goldhammer, F. (2005). The structure of the relationship between attention and intelligence. *Intelligence*, *33*(6), 589–611. https://doi.org/10.1016/j.intell.2005.07.001

Sheng, Y., & Sheng, Z. (2012). Is Coefficient Alpha Robust to Non-Normal Data? *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00034

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics*, *77*(6), 1863–1880. https://doi.org/10.3758/s13414-015-0899-0

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working Memory Capacity and Fluid Intelligence: Maintenance and Disengagement. *Perspectives on Psychological Science*, *11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Shipstead, Z., Lindsey, D. R. B., 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*, 116–141. https://doi.org/10.1016/j.jml.2014.01.004

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 word solution. *Dialogue: The Official Newsletter of the Society for Personality and Social Psychology*, *26*, 4–7.

Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143*(2), 850–886. https://doi.org/10.1037/a0033981

Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence*, *49*, 110–128. https://doi.org/10.1016/j.intell.2015.01.005

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

Unsworth, N., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica*, *150*, 14–25. https://doi.org/10.1016/j.actpsy.2014.04.001

Unsworth, N., Miller, J. D., Lakey, C. E., Young, D. L., Meeks, J. T., Campbell, W. K., & Goodie, A. S. (2009). Exploring the Relations Among Executive Functions, Fluid Intelligence, and Personality. *Journal of Individual Differences*, *30*(4), 194–200. https://doi.org/10.1027/1614-0001.30.4.194

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, *38*(1), 111–122. https://doi.org/10.1016/j.intell.2009.08.002

Unsworth, N., Redick, T. S., Spillers, G. J., & Brewer, G. A. (2012). Variation in working memory capacity and cognitive control: Goal maintenance and microadjustments of control. *Quarterly Journal of Experimental Psychology*, *65*(2), 326–355. https://doi.org/10.1080/17470218.2011.597865

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., & 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., Spillers, G. J., & Brewer, G. A. (2009). Examining the relations among working memory capacity, attention control, and fluid intelligence from a dual-component framework. *Psychology Science Quarterly*, *51*(4), 388–402.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Dynamics of context-dependent recall: An examination of internal and external context change. *Journal of Memory and Language*, *66*(1), 1–16. https://doi.org/10.1016/j.jml.2011.05.001

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence*, *35*(5), 427–449. https://doi.org/10.1016/j.intell.2006.09.001

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*(11), 418–424. https://doi.org/10.1016/j.tics.2008.07.005

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. *Psychological Review*, *115*(2), 518–525. https://doi.org/10.1037/0033-295X.115.2.518

Verguts, T., & Notebaert, W. (2009). Adaptation by binding: a learning account of cognitive control. *Trends in Cognitive Sciences*, *13*(6), 252–257. https://doi.org/10.1016/j.tics.2009.02.007

von Bastian, C. C., & Druey, M. D. (2017). Shifting Between Mental Sets: An Individual Differences Approach to Commonalities and Differences of Task Switching Components. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0000333 von Bastian, C. C., & Eschen, A. (2016). Does working memory training have to be adaptive? *Psychological Research*, *80*(2), 181–194. https://doi.org/10.1007/s00426-015-0655-z

von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*(1), 108–115. https://doi.org/10.3758/s13428-012-0224-y

von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, *145*(2), 246–258. https://doi.org/10.1037/xge0000120

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Wiecki, T. V., & Frank, M. J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. *Psychological Review*, *120*(2), 329–355. https://doi.org/10.1037/a0031542

Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology*, *18*(1), 18–50. https://doi.org/10.1080/09541440500215921

Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959.

Zheng, Y., Myerson, J., & Hale, S. (2000). Age and individual differences in visuospatial processing speed: Testing the magnification hypothesis. *Psychonomic Bulletin & Review*, *7*(1), 113–120. https://doi.org/10.3758/BF03210729

Footnotes

[1]Instructions at the beginning of the time estimation task were misunderstood by 17 participants. To keep them in the final sample, we removed the first block from the analyses for these participants.

[2]That is, *go* RTs were rank ordered and the SSRT was computed by subtracting the mean stop-signal delay from the *go* RT that corresponded to the probability of inhibiting the response (see Schachar et al., 2007).

[3]In these models, executive control – in particular, inhibition – was divided into a top-down biasing of (prefrontal) goal-related representations and a more direct inhibitory process on response selection (via subcortical areas). In Herd et al. (2014), these processes were assessed with a Stroop task and a task-switching paradigm, respectively. In Wiecki and Frank (2013), the first process was measured with the stop-signal task. The second process was assessed with three tasks (i.e., an antisaccade task, a saccade-overriding task, and a Simon task), but only a single task representation of the process is reported, making it impossible to determine whether the process parameters differed between the tasks and whether they correlated across the tasks.

[4]One could hold against this possibility the finding that individuals with faster prosaccades committed more antisaccade errors (Crawford, Parker, Solis-Trapala, & Mayes, 2011; Schaeffer et al., 2015) so that it appears unlikely that high antisaccade error rates reflect slow processing speed. This correlation, however, is well explained by a race model of the antisaccade task, according to which a prosaccade and an antisaccade are programmed in parallel, racing towards a criterion: Faster programming of prosaccades implies that they more often win the race, leading to errors in the antisaccade condition. It could still be the case that individuals who are faster at extracting information from perceptual stimuli – rather than faster at

programming reflexive saccades – do better on the antisaccade task, and it is the speed of extracting stimulus information that is correlated with fluid intelligence.

Table 1

Correlation Coefficients Between Latent Variables of Executive Control, Working Memory Capacity (WMC) and Fluid Intelligence (gF) from Previous Studies.

Study	Executive-control label	WMC label	Executive control & WMC	Executive control & gF	WMC & gF	Problems with the executive-control factor (if present)
Brydges et al. (2012)	inhibition	working memory	All executive control and WMC tasks were forced to load on a single factor labelled executive functioning.		.89*	
Chuderski (2014)	antisaccade tasks	complex spans	.64*	.55[a]	.66*	only antisaccade tasks[b]
		relational integration	.64*		.71*	
Chuderski (2015a)	executive control	binding in working memory	.60	n.a.	n.a.	predominance of the antisaccade task
		storage capacity	.82	n.a.	n.a.	
Chuderski & Necka (2012, exp. 4)	executive control	focus of attention	.32*	n.a.	n.a.	n-back tasks used to derive measured of both executive control and WMC + predominance of the 5-back task
Chuderski & Necka (2012, exp. 5)	executive control	focus of attention	.13	n.a.	n.a.	n-back tasks used to derive measured of both executive control and WMC

Study	Executive-control label	WMC label	Executive control & WMC	Executive control & gF	WMC & gF	Problems with the executive-control factor (if present)
						+ predominance of the 4-back task
Chuderski et al. (2012, exp. 1)	attention control	storage capacity	.63*	.43*	.87*	
	interference resolution		-.01	n.a.	n.a.	predominance of the picture-word interference task
	response inhibition		.24*	n.a.	n.a.	predominance of the go/no-go task
Chuderski et al. (2012, exp. 2)	attention control	scope of attention	.87*	n.a.	n.a.	predominance of the antisaccade task
		relational integration	.60*	n.a.	n.a.	
Kane et al. (2016)	attention restraint	WMC	-.64*	n.a.	n.a.	predominance of the antisaccade task
	attention constrain		-.40*	n.a.	n.a.	high error variances
	coefficient of variation (based on attention control tasks)[c]		-.32*	n.a.	n.a.	
Keye et al. (2013)[d]	horizontal Simon	working memory	-.02	n.a.	n.a.	low factor loadings
	vertical Simon		-.06	n.a.	n.a.	
Keye et al. (2010)[d]	Flanker	working memory	-.02	n.a.	n.a.	high error variances
	Simon		-.17	n.a.	n.a.	

Study	Executive-control label	WMC label	Executive control & WMC	Executive control & gF	WMC & gF	Problems with the executive-control factor (if present)
Klauer et al. (2010, exp. 1)[e]	inhibition	working memory	.58*	n.a.	n.a.	
Klauer et al. (2010, exp. 2)	inhibition	working memory	.37	n.a.	n.a.	
McCabe et al. (2010)[f]	executive functioning	WMC	.97*	n.a.	n.a.	
McVay & Kane (2012)	attention control	WMC	.73*	n.a.	n.a.	
Pettigrew & Martin (2014)	interference resolution	working memory	-.13	n.a.	n.a.	predominance of the Brown-Peterson task
Redick et al. (2016)	attention control	WMC	.72*	n.a.	n.a.	
Schweizer & Moosbrugger (2004)[g]	attention	working memory	.50* .50*	.36* .50*	.54* .22	loadings were forced to be equal across the measures
Schweizer et al. (2005)	attention[h]	(not measured)	n.a.	.57*	n.a.	
Shipstead et al. (2014)	attention control	running memory	.09	.69*	.43*	predominance of the antisaccade task
		complex spans	.68*		.71*	
		visual arrays	.74*		.68*	
		visual arrays + filter	.41*		.16	
Unsworth (2015, exp. 1)	coefficient of variation based	WMC	-.30*	-.30*	.53*	Coefficient of variation was computed across all trials

Study	Executive-control label	WMC label	Executive control & WMC	Executive control & gF	WMC & gF	Problems with the executive-control factor (if present)
	on executive-control tasks[i]					(incongruent, congruent and neutral)
Unsworth (2015, exp. 2)	coefficient of variation based on attention control tasks[i]	WMC	-.47*	n.a.	n.a.	Coefficient of variation was computed across all trials (incongruent, congruent and neutral)
Unsworth (2015, exp. 3)	coefficient of variation based on attention control tasks[i]	WMC	-.41*	-.68*	.67*	Coefficient of variation was computed across all trials (incongruent, congruent and neutral)
Unsworth & McMillan (2014)	attention control	WMC	.62*	.78*	.67*	
Unsworth, Miller et al. (2009)	response inhibition	working memory	.35*	.76*	.40*	predominance of the antisaccade task
Unsworth et al. (2014)	attention control	storage capacity processing	.51* .82* -.53*	.77*	.57* .71* -.53*	
Unsworth et al. (2010)	response inhibition	WMC	.30*	.55*	.79*	
Unsworth & Spillers (2010)	attention control	WMC	.58*	.45*	.53*	
Unsworth, Spillers et al. (2009)	attention control	WMC	.41*	.70*	.66*	little evidence in a data re-analysis[j]

Note. As executive control and WMC latent variables were labelled differently across studies, the second and third column refer to the factor names used in the studies and list the different factors if different factors were computed for the executive control and WM constructs. This was not necessary for gF because in all studies, it was referred to as either fluid intelligence or reasoning. If a model included a path from one construct to a second without any other predictor to the second construct, this path coefficient is equivalent to a correlation, and we report it in the table. In any other case, the path coefficient does not reflect the bivariate correlation and therefore we do not report it. This leads to the exclusion of some studies (Chuderski, 2015b; Dang, Braeken, Colom, Ferrer, & Liu, 2014; Shipstead, Harrison, & Engle, 2015, 2016). Moreover, as conventional measures of updating are confounded with WMC (e.g., Ecker et al., 2010), we do not include estimates of the correlation between updating and WMC or gf (Friedman et al., 2006, 2008; Hedden & Yoon, 2006; Hull et al., 2008; Miyake et al., 2000; van der Sluis et al., 2007). A factor loading was considered as predominant if the highest or the two highest loadings were 1.5 times larger than the subsequent one. As this criterion is arbitrary, we present all factor loadings in Appendix B.

n.a. = not available.

* $p < .05$.

[a]No information was reported about the significance of this correlation.

[b]Only antisaccade tasks loaded on the executive-control factor. Two Stroop tasks and one stop-signal task were also measured but not reported.

[c]For each executive-control measure, a coefficient of variation (i.e., standard deviation / mean RT) was computed across either congruent or neutral trials. Then, these measures were used to create a factor called “coefficient of variation”.

[d]Instead of computing the difference between incongruent and congruent trials as the dependent measure, executive control was

measured in a nested model on which all measures loaded on a general performance factor while the measures reflecting RTs on incongruent trials had free loadings on the executive-control factor and the measures reflecting RTs on congruent trials had their loadings on the executive-control factor fixed to zero.

[e]Although the model with distinct executive control and WMC factors had a good fit, Klauer et al. (2010) opted for the subsequent analyses for a more parsimonious model in which both latent variables were merged and which fitted as well as the 2-factor model.

[f]In that study, the age of the sample ranged from 18 to 98 years.

[g]Schweizer and Moosbrugger (2004) computed models in which gF was measured either with the Raven Advanced Progressive Matrices (RAPM) or the "Zahlen-Verbindungs-Test" (ZVT). Values in the first row of the cell refers to the model in which gF was measured with the RAPM; values in the second row of the cell refer to the model in which gF was measured with the ZVT.

[h]Attention refers to a common higher-level latent variable, which subsumed two latent variables (i.e., perceptual processing and higher mental/executive-control processing).

[i]In that study, the data from Unsworth and Spillers (2010), Unsworth et al. (2012), and Unsworth and McMillan (2014) were reanalyzed as Experiments 1, 2, and 3, respectively. For the executive-control measures, the focus was on the intra-individual variability, computed as a coefficient of variation (i.e., standard deviation / mean RT). Thus, larger values for these executive control-measures indicate worse abilities, whereas larger values for the WMC and gF measures indicate better abilities. This explained why the correlations involving the executive-control measures were negative.

[j]Gignac and Kretzschmar (2016) reanalyzed the data using a bifactor model approach and found little evidence for the existence of the

executive-control latent variable.

Table 2

Sample Characteristics and Background Measures.

Measure	Sample
Sample size	181
Age (years)	21.3 (2.2)
Age range	18-28
Gender (female/male)	139/42
Education (years)	13 (2.8)
Educational level[a]	5.2 (0.7)
BDI-II score[b]	9.6 (7.6)

Note. Standard deviations are given in parentheses. BDI-II = Beck Depression Inventory II (Hautzinger et al., 2006).

[a]Education level ranged from 1 (no or less than 9 school years) to 8 (PhD).

[b]Depression score ranged from 0 (minimal depression) to 63 (severe depression).

Table 3

Descriptive Statistics.

Construct/Task	*M*	*SD*	Min.	Max.	Skew	Kurtosis	Reliability
Executive Control							
Color Stroop	8.43	8.64	-8.33	39.58	0.74	0.61	.58
Number Stroop	20.98	9.63	-6.25	46.88	0.16	0.03	.62
Arrow flanker	39.79	14.50	-1.04	80.21	-0.07	-0.13	.85
Letter flanker	8.37	6.59	-9.37	28.13	0.08	-0.15	.26
Simon	20.65	13.75	-15.63	59.38	0.04	-0.19	.83
Antisaccade	22.88	19.78	-24.17	73.96	0.15	-0.59	.91
Stop-signal	39.81	19.56	7.29	100.00	0.74	0.11	.88
Fluid Intelligence							
Letter sets	27.33	13.78	3.33	76.67	1.10	1.45	.77
Locations	45.74	16.18	7.14	96.43	0.04	-0.09	.72
RAPM	41.62	20.98	0.00	100.00	0.33	-0.19	.63
Relationships	29.24	14.61	3.33	80.00	0.93	0.99	.75
Syllogisms	39.47	12.46	3.33	73.33	-0.01	0.20	.68
Working Memory Capacity							
Numerical updating	45.38	20.58	1.19	84.52	-0.30	-0.81	.95
Spatial updating	33.09	14.38	8.33	84.72	0.97	0.85	.98
Numerical complex span	59.02	16.06	12.17	97.83	-0.08	-0.29	.80
Spatial complex span	55.43	18.36	6.25	100.00	-0.01	-0.56	.86
Time Estimation							
With timer[a]	1.25	0.01	1.23	1.28	0.06	-0.12	.86
Without timer[a]	1.87	0.03	1.78	1.96	-0.18	-0.20	.94

Note. Scores were computed as the difference in error rates (in %) between incongruent and congruent trials for the color Stroop, number Stroop, arrow flanker, letter flanker, and Simon task, as error rates in antisaccade trials for the antisaccade task, as error rates in stop trials for the stop-signal task, as error rates in the fluid intelligence and working memory capacity tasks, and as response time precision (i.e., the absolute difference between the stimulus offset time and the RT) for the time estimation tasks. Reliabilities were calculated by adjusting split-half correlations with the Spearman–Brown prophecy formula. Split-half correlations were computed between odd and even items, except in the stop-signal task in which they were computed between the first two blocks and the last two blocks. Min. = minimum; Max. = maximum.

[a]A Yeo-Johnson transformation (Yeo & Johnson, 2000) was applied on the scores because skew and kurtosis were smaller than -3 or larger than 3.

Table 4

Pearson Correlation Coefficients.

	Color Stroop	Number Stroop	Arrow flanker	Letter flanker	Simon	Antisaccade	Stop-signal
Number Stroop	*.11*	-					
	[-.03, .25]						
Arrow flanker	*.04*	*-.01*	-				
	[-.12, .19]	[-.15, .13]					
Letter flanker	*.08*	*.04*	*.09*	-			
	[-.06, .22]	[-.10, .18]	[-.06, .24]				
Simon	*-.11*	*-.11*	*.07*	*-.002*	-		
	[-.26, .03]	[-.24, .03]	[-.09, .23]	[-.15, .14]			
Antisaccade	*-.05*	*-.01*	*-.04*	*-.01*	-.16*	-	
	[-.18, .09]	[-.14, .13]	[-.19, .10]	[-.17, .14]	[-.31, -.005]		
Stop-signal	*-.04*	*.05*	*-.10*	*-.09*	.15	*.05*	-
	[-.16, .08]	[-.09, .19]	[-.25, .05]	[-.23, .04]	[.002, .29]	[-.09, .20]	
Letter sets	*-.06*	*.03*	*-.09*	**-.22***	*.04*	*.01*	*-.04*
	[-.22, .10]	[-.10, .16]	[-.28, .09]	[-.35, -.08]	[-.13, .21]	[-.13, .15]	[-.18, .11]
Locations	*-.08*	*-.09*	*.09*	*-.08*	*.06*	*-.03*	-.16*
	[-.21, .06]	[-.22, .03]	[-.06, .23]	[-.22, .06]	[-.10, .21]	[-.18, .11]	[-.31, -.01]
RAPM	*.02*	*-.05*	*-.06*	-.17*	.15*	*.06*	*-.03*
	[-.13, .16]	[-.19, .10]	[-.22, .09]	[-.30, -.04]	[-.02, .32]	[-.09, .21]	[-.19, .13]
Relationships	*-.08*	*.01*	-.14	*-.13*	*.03*	*.05*	*.01*
	[-.24, .08]	[-.12, .14]	[-.34, .05]	[-.27, .01]	[-.14, .19]	[-.09, .20]	[-.14, .15]
Syllogisms	*-.02*	*-.02*	*.02*	*.05*	*.01*	*-.02*	*-.09*
	[-.14, .11]	[-.16, .12]	[-.11, .15]	[-.10, .19]	[-.13, .15]	[-.16, .13]	[-.23, .06]
Numerical updating	*.01*	*-.03*	*.07*	*-.03*	*.03*	*.05*	*-.05*
	[-.14, .17]	[-.16, .11]	[-.08, .23]	[-.18, .11]	[-.11, .17]	[-.10, .20]	[-.20, .11]
Spatial updating	*.01*	*-.01*	*.04*	-.15*	.20*	*-.04*	*.04*
	[-.14, .18]	[-.15, .14]	[-.13, .21]	[-.30, -.002]	[.04, .38]	[-.18, .10]	[-.08, .16]

	Color Stroop	Number Stroop	Arrow flanker	Letter flanker	Simon	Antisaccade	Stop-signal
Numerical complex span	*.11*	*.11*	*.09*	*-.08*	*.02*	*.06*	*-.05*
	[-.03, .25]	[-.04, .25]	[-.07, .24]	[-.23, .07]	[-.14, .19]	[-.10, .23]	[-.23, .12]
Spatial complex span	*.11*	*.06*	*.04*	*-.08*	*.06*	*.001*	.14
	[-.03, .25]	[-.07, .20]	[-.13, .20]	[-.23, .07]	[-.10, .21]	[-.15, .16]	[.002, .29]
Time estimation with timer	*-.02*	*-.05*	-.19*	*-.005*	.14	*-.11*	*.01*
	[-.16, .12]	[-.20, .09]	[-.34, -.05]	[-.13, .12]	[-.02, .30]	[-.24, .02]	[-.13, .15]
Time estimation without timer	*.10*	*.11*	*-.01*	*.02*	*-.01*	*-.06*	*.12*
	[-.04, .24]	[-.04, .26]	[-.17, .15]	[-.12, .16]	[-.16, .13]	[-.20, .08]	[.001, .25]

(Table 4 continues)

	Letter sets	Locations	RAPM	Relation ships	Syllo -gisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.s.	Timer
Locations	**.38***	-								
	[.24, .52]									
RAPM	**.44***	**.33***	-							
	[.32, .57]	[.20, .46]								
Relationships	**.55***	**.40***	**.33***	-						
	[.42, .69]	[.27, .53]	[.19, .47]							
Syllogisms	.16*	**.33***	*.10*	**.31***	-					
	[.03, .30]	[.21, .45]	[-.05, .25]	[.18, .43]						
Num. upd.	**.44***	**.31***	**.33***	**.34***	**.19***	-				
	[.33, .55]	[.18, .45]	[.21, .46]	[.21, .47]	[.04, .35]					
Spatial upd.	**.44***	**.39***	**.28***	**.49***	**.30***	**.46***	-			
	[.31, .57]	[.27, .52]	[.14, .43]	[.36, .62]	[.17, .43]	[.34, .57]				
Num. c.s.	**.25***	*.12*	**.23***	.20*	*-.01*	**.41***	**.29***	-		
	[.12, .38]	[-.03, .28]	[.10, .37]	[.06, .33]	[-.16, .14]	[.28, .55]	[.14, .44]			

	Letter sets	Locations	RAPM	Relation ships	Syllo -gisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.s.	Timer
Spatial	**.37***	**.30***	**.29***	**.33***	*.10*	**.49***	**.48***	**.49***	-	
c.s.	[.23, .50]	[.17, .44]	[.15, .43]	[.19, .46]	[-.05, .24]	[.38, .61]	[.36, .60]	[.38, .61]		
Time estimation	*.04*	*.07*	.14	*.05*	*-.09*	.16*	*.08*	*.03*	.14	-
with timer	[-.12, .21]	[-.06, .20]	[.01, .28]	[-.10, .19]	[-.23, .06]	[.01, .32]	[-.07, .23]	[-.11, .18]	[.001, .29]	
Time estimation	.17*	**.28***	.18*	**.21***	*.12*	**.30***	**.27***	.20*	**.44***	*.05*
without timer	[.03, .32]	[.16, .40]	[.06, .31]	[.07, .34]	[-.03, .27]	[.18, .42]	[.14, .40]	[.07, .33]	[.32, .56]	[-.11, .20]

Note. Ninety-five % bootstrapped confidence intervals (10000 random samples) are presented in brackets. Correlations for which the Bayes factor suggested positive to strong evidence for the alternative hypothesis (BF_{10}) are presented in bold; correlations for which the Bayes factor suggested positive to strong evidence for the null hypothesis (BF_{01}) are presented in italics. Bayes factors for each correlation are presented in Appendix G. Num. = Numerical; upd. = updating; c.s. = complex span.

* $p < .05$.

Table 5

Goodness-Of-Fit Statistics and Model Comparisons.

Construct/Model	χ^2	df	p	CFI	RMSEA [90% CI]	SRMR	AIC	BIC
Fluid Intelligence and Working Memory Capacity								
1	*53.66*	*26*	*.001*	*.93*	*.08 [.05, .11]*	*.06*	*13309.04*	*13369.81*
2	126.12	27	<.001	.77	.14 [.12, .17]	.21	13379.50	13437.08
3	79.28	27	<.001	.88	.10 [.08, .13]	.07	13332.66	13390.24
Executive Control								
6	*0.16*	*2*	*.923*	*1*	*0 [0, .05]*	*.01*	*5968.44*	*5994.02*
7 (saturated model)	*0*	*0*	*NA*	*0*	*0 [0, 0]*	*0*	*4485.96*	*4505.15*

Note. Models with acceptable fit statistics are presented in italics. Model 1 = Two-factor model with fluid intelligence (gF) and working memory capacity (WMC) as correlated but distinct factors; Model 2 = Two-factor model with gF and WMC as uncorrelated factors; Model 3 = Single-factor model in which all gF and WMC tasks loaded on a factor; Model 6 = Single-factor model in which only the tasks used by Friedman an Miyake (2004) were included (i.e., the antisaccade, stop-signal, color Stroop and arrow flanker tasks); Model 7 = Single-factor model in which only the tasks used by Miyake et al. (2000) were included (i.e., the antisaccade, stop-signal, and color Stroop tasks). CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

Table 6

Standardized factor loadings and error variances.

Model / Task	Factor Loadings			Error Variances		
	Estimate	CI	*p*	Estimate	CI	*P*
Model 1						
Letter sets	.74	[.60, .88]	< .001	.45	[.31, .58]	< .001
Locations	.57	[.42, .71]	< .001	.68	[.52, .83]	< .001
RAPM	.54	[.39, .69]	< .001	.71	[.54, .86]	< .001
Relationships	.71	[.56, .85]	< .001	.50	[.36, .63]	< .001
Syllogisms	.33	[.18, .49]	< .001	.89	[.70, 1.07]	< .001
Numerical updating	.70	[.56, .84]	< .001	.51	[.37, .64]	< .001
Spatial updating	.69	[.55, .83]	< .001	.52	[.38, .66]	< .001
Numerical complex span	.53	[.38, .68]	< .001	.72	[.55, .88]	< .001
Spatial complex span	.71	[.57, .85]	< .001	.52	[.36, .63]	< .001
Model 6						
Antisaccade	-.17	[-.17, .51]	.340	.97	[.74, 1.19]	< .001
Stop-signal	-.35	[-.21, .90]	.226	.88	[.46, 1.29]	< .001
Color Stroop	.14	[-.47, .19]	.421	.98	[.76, 1.19]	< .001
Arrow flanker	.29	[-.76, .19]	.238	.92	[.59, 1.23]	< .001
Model 7						
Antisaccade	.26	[-.43, .95]	.464	.93	[.53, 1.33]	< .001
Stop-signal	.21	[-.36, .78]	.469	.96	[.65, 1.25]	< .001
Color Stroop	-.19	[-.69, .32]	.473	.97	[.69, 1.23]	< .001

Note. Model 1 = Two-factor model with fluid intelligence (gF) and working memory capacity (WMC) as correlated but distinct factors; Model 6 = Single-factor model in which only the tasks used by Friedman an Miyake (2004) were included (i.e., the antisaccade, stop-signal, color Stroop and arrow flanker tasks); Model 7 = Single-factor model in which only the tasks used by Miyake et al. (2000) were included (i.e., the antisaccade, stop-signal, and color Stroop tasks). Ninety-five % CI = confidence interval; RAPM = Raven's Advanced Progressive Matrices.

Table 7

Relations Between the Individual Measures of Executive Control and the Latent Variables of Fluid Intelligence and Working Memory Capacity: Goodness-Of-Fit Statistics and Regression Coefficients.

Task	χ^2	*p*	CFI	RMSEA [90% CI]	SRMR	AIC	BIC	Regression to gF			Regression to WMC		
								coefficient	*p*	BF_{01}	coefficient	*p*	BF_{01}
Color Stroop	56.90	.006	.94	.06 [.03-.09]	.06	14601.35	14668.52	-.08	.325	8.28	.08	.318	8.20
Number Stroop	59.58	.003	.94	.07 [.04-.09]	.06	14644.84	14712.01	-.02	.814	13.09	.04	.646	12.12
Arrow flanker	62.45	.001	.93	.07 [.04-.10]	.06	14788.08	14855.25	-.10	.227	6.52	.08	.345	8.61
Simon	65.39	.001	.92	.07 [.05-.10]	.06	14772.24	14839.41	.08	.319	8.21	.13	.131	4.40
Antisaccade	57.55	.005	.94	.06 [.03-.09]	.06	14905.97	14973.13	.03	.726	12.65	.02	.816	13.01
Stop-signal	70.32	< .001	.91	.08 [.05-.10]	.06	14899.49	14966.66	-.08	.370	9.03	.05	.565	11.44

Note. Each executive-control measure was the predictor of the gF and WMC factors (*df*s = 33 for the χ^2 statistic). All models had acceptable fit statistics. The Bayes Factor (BF_{01}) was computed by using the difference between the BIC for the alternative hypothesis (i.e., the 2-factor model in which the regressions to both gF and WMC were freely estimated) and the BIC for the null hypothesis (e.g., the 2-factor model in which the regression to either gF or WMC, respectively, was set to 0). CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion; gF = fluid intelligence; WMC = working memory capacity.

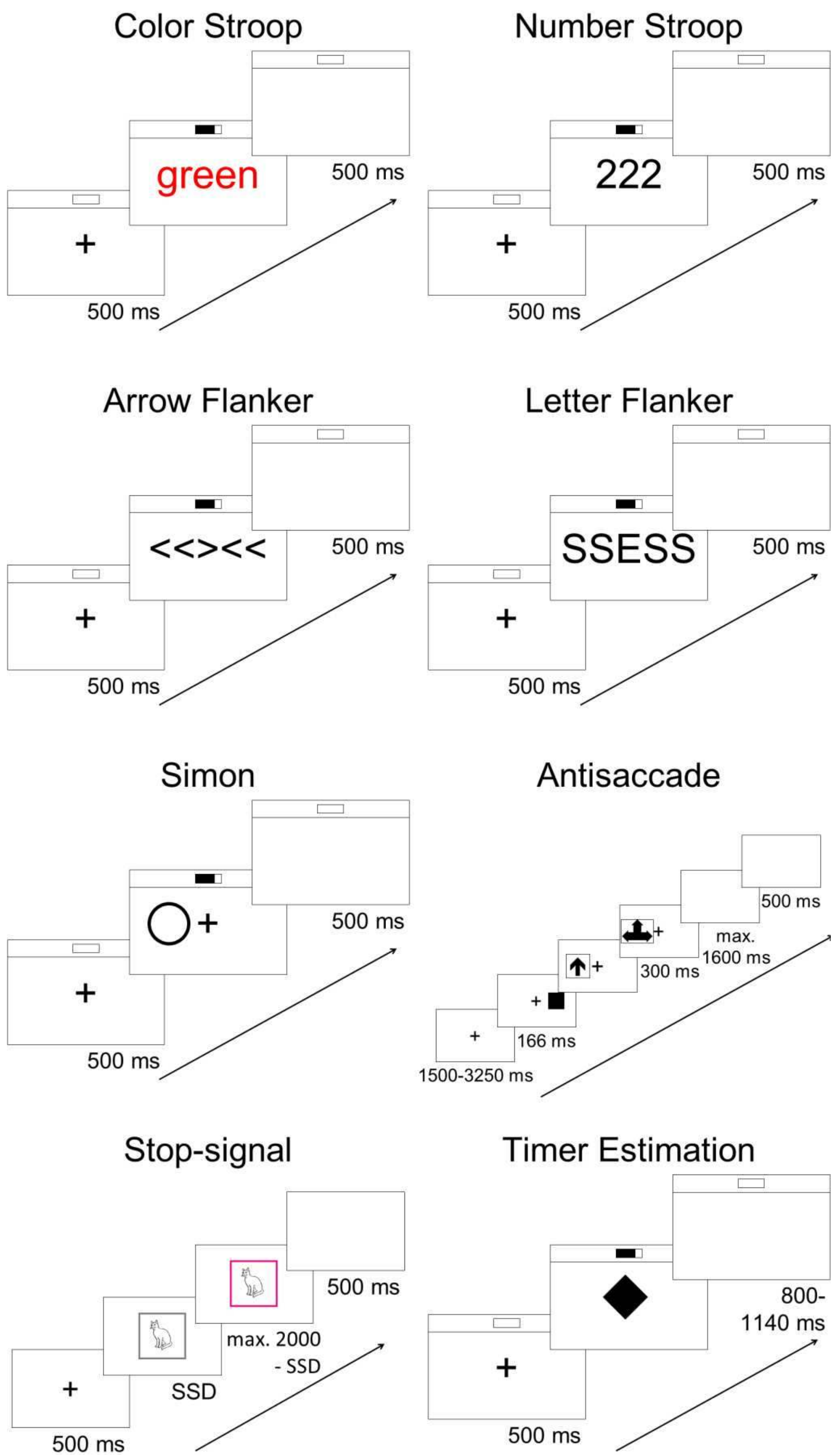

Figure 1. Example of one trial sequence from the experimental blocks for each executive-control task and the time estimation task (SSD = stop-signal delay).

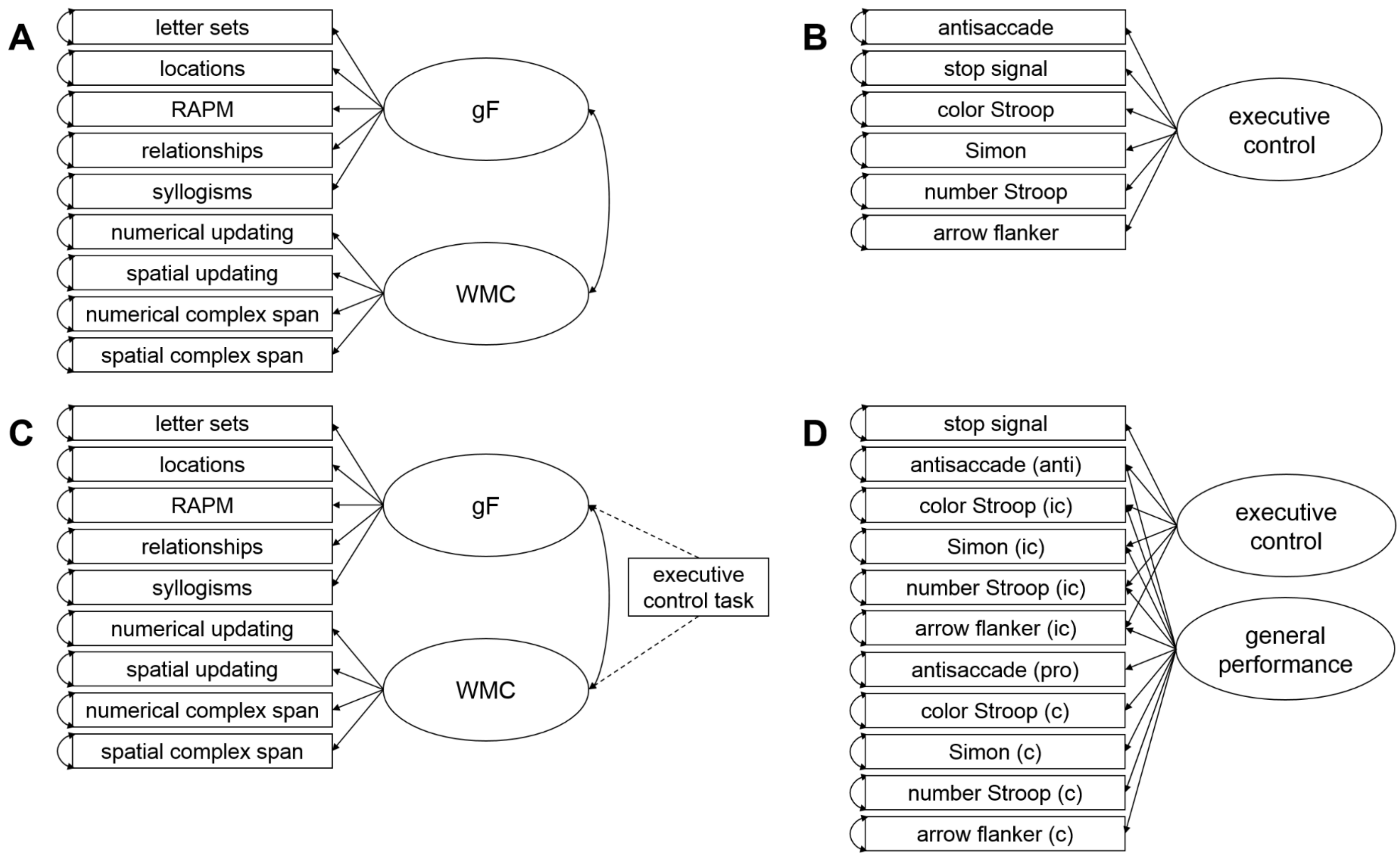

Figure 2. (A) Two-factor model with fluid intelligence (gF) and working memory capacity (WMC) as correlated but distinct factors (Model 1). (B) One-factor model in which all tasks assumed to assess executive control loaded a single latent variable (Model 4). (C) Example of Model 8 used to determine the regression coefficients (dotted arrows) between each individual measure of executive

control and the latent variables of gF and WMC. (D) Bi-factor model in which antisaccade (anti) and prosaccade (pro) trials of the antisaccade task and incongruent (ic) and congruent (c) trials of each task were forced to load on a general performance factor, and antisaccade and incongruent trials as well as the stop-signal task were forced to load on an executive-control factor. RAPM = Raven's Advanced Progressive Matrices.

Appendix A

Tasks Used to Measure Executive Control

Table A1

Description of Tasks Typically Used to Measure Executive Control.

Task	Description	Studies (examples)
Antisaccade*	Participants were asked to identify a stimulus which is presented very briefly on the side opposite of a flashing cue. Thus, to perform this task successfully, participants had to suppress the reflexive saccade to the cue and perform a saccade in the opposite direction to identify the stimulus. The stimulus to identify could be an arrow, a letter or a digit.	Chuderski (2014, 2015b); Chuderski et al. (2012); Friedman & Miyake (2004); Friedman et al. (2006, 2008); Klauer et al. (2010); McVay & Kane (2012); Miyake et al., (2000); Shipstead et al. (2014); Unsworth (2015); Unsworth & McMillan (2014); Unsworth, Miller, et al. (2009); Unsworth et al. (2014); Unsworth & Spillers (2010); Unsworth, Spillers, et al. (2009)
Brown-Peterson	Participants learned and later freely recalled successive lists that were composed of words drawn from the same category.	Friedman & Miyake (2004); Pettigrew & Martin (2014)
Color Stroop*	Participants saw a color word printed in an incongruent color (i.e., the word “red” printed in green). They were asked to name the color of the word and to inhibit its meaning.	Brydges et al. (2012); Chuderski et al. (2012); Friedman & Miyake (2004); Friedman et al. (2006, 2008); Klauer et al. (2010); Miyake et al., (2000); Shipstead et al. (2014); Unsworth (2015); Unsworth & McMillan (2014); Unsworth & Spillers (2010)
Compatibility reaction time	Participants were asked to press one of two buttons depending on whether the lengths of two lines were the same or different. A first series of blocks built up a prepotent response for the buttons required for same and	Brydges et al. (2012)

Task	Description	Studies (examples)
	different lengths, after which a final block was administered, where the required button for a response were swapped.	
Cued recall	Participants saw either one or two lists of four words each and had to retrieve the word on the most recent list that belongs to a cued category, ignoring any previous lists.	Friedman & Miyake (2004); Pettigrew & Martin (2014)
Cued search / cued flanker	Participants were asked to report via key-press the direction of a target letter (normal vs. mirror-reversed). The letter was presented equally often in one of several locations, and the possible target locations were precued. Non-target locations were populated by irrelevant letters and one lure. The presence of the lure required participants to focus on the cued locations only.	Kane et al. (2016); Redick et al. (2016)
Disengage	In a first phase, participants were presented with a red square frame with a gap along with three more differently colored square frames. After some time, all square frames were masked with color patches, and participants were asked to report the direction of the gap. In a second phase ("attentional disengagement"), participants performed the same task, except that on 1/3 of the trials, a colored square frame (distractor) was briefly presented prior to the target onset.	Unsworth et al. (2014)
Flanker*	Participants were asked to identify the central character while ignoring the flanking characters. The stimulus to identify could be an arrow or a letter ("arrow flanker task" and "letter flanker task", respectively). A response deadline is sometimes given ("conditional accuracy flanker task"). In the "circle" version of the task, targets appeared in one	Friedman & Miyake (2004); Kane et al. (2016); Keye et al. (2009, 2010); Klauer et al. (2010); Shipstead et al. (2014); Unsworth (2015); Unsworth & McMillan (2014); Unsworth, Miller, et al. (2009); Unsworth & Spillers (2010); Unsworth, Spillers, et al. (2009)

Task	Description	Studies (examples)
	of several locations in a circular arrangement, and distractors appeared in one position clockwise and counterclockwise from the target; the other positions were occupied by irrelevant symbols. In the "masked" version, the stimulus array (i.e., target and flankers) was presented above or below a fixation cross for very short time (50 and 70 ms) and then was masked.	
Fluency	Participants were given 1 min to generate as many words as possible for a given letter.	McCabe et al. (2010); Shipstead et al. (2016)
Frankfurt Adaptive Concentration-Performance Test	Participants were asked to discriminate between target and non-target items. The items were geometrical figures that differed with respect to shape (square or circle) and the number of dots included (two or three dots).	Schweizer & Moosbrugger (2004); Schweizer et al. (2005)
Go/no-go	Participants were asked to press the right button as soon as possible when a stimulus appeared on screen (go trials), except when an "X" was presented, in which case they should withhold a response (no-go trials).	Brydges et al. (2012); Chuderski et al. (2012); McVay & Kane (2012); Unsworth (2015)
Mental arithmetic	Participants were asked to complete a series of progressively more difficult arithmetic problems that were verbally spoken and had to be computed without aid of pen and paper; the answer was given verbally.	McCabe et al. (2010)
N-back	Participants were asked to match the current item in a continuous stream of stimuli with an item that occurred n items ago.	Chuderski and Necka (2012)
Number Stroop*	Participants were asked to report the number of digits in a row while ignoring to read the digit value.	Chuderski et al. (2012); McVay & Kane (2012)

Task	Description	Studies (examples)
Picture-word Stroop / picture-word interference	Participants saw a picture with a superimposed distracter word and they were asked to respond by naming the picture (while ignoring the word).	Chuderski et al. (2012)
Psychomotor vigilance	Participants were presented with a row of zeros and after a variable amount of time the zeros begin to count up in 1 ms intervals from 0 ms. Participants were asked to press the spacebar once the numbers started counting up.	Unsworth (2015); Unsworth & McMillan (2014); Unsworth & Spillers (2010); Unsworth, Spillers, et al. (2009)
Recent negatives	Participants heard a list of three words followed by a probe word and indicated whether the probe word was in the previously heard list. On recent-negative trials, the probe matched a list element from a recent trial, but not the current trial. On non-recent negative trials, the probe did not match words from the current or any recent trial. The RT difference between recent and non-recent negative probes reflects the ability to inhibit the misleading familiarity of probes matching a list element in a recent trial.	Pettigrew & Martin (2014)
Simon*	Participants are asked to respond to non-spatial features of a stimulus using manual responses. In the horizontal version, stimuli are presented at either a left or a right position, which can be congruent or incongruent with the position (left or right) of the response keys on the keyboard. In the vertical version, stimuli are presented above or below a fixation cross and response keys are arranged vertically.	Keye et al. (2009, 2010, 2013); Klauer et al. (2010)
Spatial Stroop	Participants were instructed to ignore the location of an arrow in order to respond to the direction pointed by the arrow.	Kane et al. (2016); Pettigrew & Martin (2014); Redick et al. (2016)

Task	Description	Studies (examples)
Stop-signal*	Participants performed an ongoing task (e.g., a word categorization) unless the stop-signal (i.e., a tone or a change in color frame) occurred. In this case, they had to withhold their responses. The time between the presentation of the stimulus and the stop signal is adapted such that participants can only stop their reaction successfully in 50% of the trials. The better participants achieve the 50% stopping criterion for longer delays of the stop signal relative to stimulus onset (reflected in shorter stop-signal reaction times, SSRT).	Chuderski et al. (2012); Friedman & Miyake (2004); Friedman et al. (2006, 2008); Klauer et al. (2010); Miyake et al., (2000);
Stroop matching	Participants were asked to respond whether the color word and its print color were the same or not.	Klauer et al. (2010)
Tower of Hanoi	The Tower of Hanoi consists of three rods and a number of disks of different sizes, which can slide onto any rod. Participants were asked to move the entire starting configuration to another rod by moving only one disk at a time and by placing no disk on the top of a smaller disk.	Deng et al. (2014)
Trail making test	Participants were asked to connect in sequential order a set of targets. In the first part, all targets were numbers. In the second part, participants alternated between numbers and letters.	Deng et al. (2014); Shipstead et al. (2016)
Wisconsin Card Sorting test	Participants were asked to sort cards on the basis of one of three dimensions (i.e., color, shape, or number). After some trials, unbeknownst to the participant, the sorting rule changed, and cards should be sorted on another dimension. Participants received feedback (correct vs. false) after every trial in order to adapt their sorting rule	McCabe et al. (2010)

Task	Description	Studies (examples)
48drop	Participants were presented with either 4 or 8 colored squares (set size 4 and set size 8, respectively) and they were asked to remember as many colors as possible. After some time, one test colored square was presented at one of the original stimulus locations, and participants were asked to indicate if it was the same color as the original stimulus presented at that location.	Unsworth et al. (2014)

Note. * Task used in the present study.

Appendix B

Factor Loadings and Error Variances for the Measures Used to Assess Executive Control in Previous Studies

Table B1

Factor Loadings and Error Variances for the Measures Used to Assess Executive Control in the Studies Listed in Table 1.

Study	Executive-control label	Measures used to assess executive control	Loadings	Error Variances
Brydges et al. (2012)	executive functioning	color Stroop (incongruent – neutral, RTs)	.46	.79
		compatibility reaction time (incongruent – congruent, RTs)	.23	.95
		letter-number sequencing (similar to Trial Making test, accuracy)	.70	.51
		backward digit span (accuracy)[a]	.47	.78
		sentence repetition (accuracy)[a]	.52	.73
		Wisconsin Card Sorting test (accuracy)	.48	.77
		verbal fluency (accuracy)	.59	.65
		letter monitoring (accuracy)[a]	.53	.72
Chuderski (2014)	antisaccade tasks	spatial antisaccades (accuracy)	.96	.08
		letter antisaccades (accuracy)	.86	.26
		digit antisaccades (accuracy)	.83	.31
Chuderski (2015a)	executive control	*antisaccade (accuracy)*	*.75*	.44
		picture-word interference (incongruent – congruent, accuracy)	.36	.87
Chuderski & Necka (2012, exp. 4)	executive control	*5-back alarm rate*	*-.81*	.34
		1-back alarm rate	-.38	.86
Chuderski & Necka (2012, exp. 5)	executive control	1-back alarm rate	-.40	.84
		3-back alarm rate	*-.87*	.24
Chuderski et al. (2012, Exp. 1)	interference resolution	*picture-word interference (ratio of RTs in*	*.76*	.42

Study	Executive-control label	Measures used to assess executive control	Loadings	Error Variances
		incongruent trials to RT in congruent trials)		
		color Stroop (ratio of RTs in incongruent trials to RT in congruent trials)	.32	.90
	attention control	picture-word interference (ratio of accuracy rates in incongruent trials to accuracy rates in congruent trials)	.42	.82
		color Stroop (ratio of accuracy rates in incongruent trials to accuracy rates in congruent trials)	.18	.97
		number Stroop (ratio of accuracy rates in incongruent trials to accuracy rates in congruent trials)	.31	.90
		antisaccade (accuracy)	.57	.68
	response inhibition	stop-signal (accuracy)	.36	.87
		no-go (accuracy)	*.95*	.10
Chuderski et al. (2012, Exp. 2)	attention control	picture-word interference (accuracy)	.27	.93
		number Stroop (accuracy)	.20	.96
		antisaccade (accuracy)	.65	.58
Kane et al. (2016)	attention restraint	*antisaccade letters (accuracy)*	*.76*	.42
		antisaccade arrows (accuracy)	*.74*	.45
		go/no-go (signal-detection measure of performance)	-.47	.78
		go/no-go (standard deviation for go trials)	.47	.78
		number Stroop (RTs of incongruent trials)	.34	.88
		spatial Stroop (incongruent vs. congruent; accuracy)[b]	.27	.93

Study	Executive-control label	Measures used to assess executive control	Loadings	Error Variances
	attention constraint	arrow flanker (incongruent vs. congruent, RTs)[b]	.21	.96
		arrow flanker (incongruent vs. neutral, RTs)[b]	.29	.92
		arrow flanker (incongruent vs. congruent, RTs)[b]	.21	.96
		arrow flanker (incongruent vs. neutral, RTs)[b]	.32	.90
		conditional accuracy flanker (incongruent vs. congruent, accuracy)[b]	.46	.79
		conditional accuracy flanker (incongruent vs. neutral, accuracy)[b]	.30	.91
		masked flanker (incongruent vs. congruent, accuracy)[b]	.46	.79
		masked flanker (incongruent vs. neutral, accuracy)[b]	.36	.87
		circle flanker (incongruent vs. neutral, RTs)[b]	.19	.96
	coefficient of variation (based on executive-control tasks)	go/no-go (coefficient of variation for correct go trials)	.36	.87
		number Stroop (coefficient of variation for congruent trials)	.42	.82
		spatial Stroop (coefficient of variation for neutral trials)	.50	.75
		arrow flanker (coefficient of variation for neutral trials)	.41	.83
		arrow flanker (coefficient of variation for congruent trials)	.41	.83

Study	Executive-control label	Measures used to assess executive control	Loadings	Error Variances
		letter flanker (coefficient of variation for neutral trials)	.41	.83
		letter flanker (coefficient of variation for congruent trials)	.42	.82
Keye et al. (2013)	horizontal Simon	incongruent after congruent with stimulus repetition (RTs)	.22	n.a.
		incongruent after congruent without stimulus repetition (RTs)	.23	n.a.
		incongruent after incongruent with stimulus repetition (RTs)	.23	n.a.
		incongruent after incongruent without stimulus repetition (RTs)	.24	n.a.
	vertical Simon	incongruent after congruent with stimulus repetition (RTs)	.21	n.a.
		incongruent after congruent without stimulus repetition (RTs)	.22	n.a.
		incongruent after incongruent with stimulus repetition (RTs)	.22	n.a.
		incongruent after incongruent without stimulus repetition (RTs)	.22	n.a.
Keye et al. (2010)	Flanker	incongruent after congruent with stimulus repetition (RTs)	.53	.91[c]
		incongruent after congruent without stimulus repetition (RTs)	.56	.97[c]
		incongruent after incongruent with stimulus repetition (RTs)	.48	.95[c]
		incongruent after incongruent without stimulus repetition (RTs)	.51	.95[c]

Study	Executive-control label	Measures used to assess executive control	Loadings	Error Variances
	Simon	incongruent after congruent with stimulus repetition (RTs)	.17	.95[c]
		incongruent after congruent without stimulus repetition (RTs)	.19	.97[c]
		incongruent after incongruent with stimulus repetition (RTs)	.25	.98[c]
		incongruent after incongruent without stimulus repetition (RTs)	.27	.97[c]
Klauer et al. (2010, exp. 1)	inhibition	stop-signal (accuracy)	.58[d]	.66
		antisaccade (accuracy)	.45[d]	.80
Klauer et al. (2010, Exp. 2)	inhibition	letter flanker (incongruent – congruent, RTs)	.24	.94
		Stroop matching (incongruent – congruent, RTs)	.35	.88
		Simon (incongruent – congruent, RTs)	.46	.79
		antisaccade (antisaccade – prosaccade, RTs)	.61	.63
McCabe et al. (2010)	executive functioning	mental arithmetic (accuracy)	.52	.73
		mental control (similar to Trial Making Test, accuracy)	.61	.63
		verbal fluency (accuracy)	.42	.82
		Wisconsin Card Sorting Test (accuracy)	.69	.52
McVay & Kane (2012)	attention control	number Stroop (RTs of incongruent trials)	-.37	.86
		go/no-go (standard deviation for go trials)	-.46	.79
		go/no-go (signal-detection measure of performance)	.37	.86
		antisaccade (accuracy)	.65	.58

Study	Executive-control label	Measures used to assess executive control	Loadings	Error Variances
Pettigrew & Martin (2014)	interference resolution	recent negatives (recent vs. non-recent, RTs)[c]	.26	.93[c]
		cued recall (accuracy)[b]	.43	.82[c]
		Brown-Peterson (accuracy)[b]	*.77*	.41[c]
		letter flanker (incongruent vs. neutral, RTs)[b]	-.09	.99[c]
		picture-word interference (incongruent vs. congruent, RTs)[b]	.20	.96[c]
		nonverbal (spatial) Stroop (incongruent vs. neutral, RTs)[b]	.32	.90[c]
		color Stroop (incongruent vs. neutral, RTs)[b]	.26	.93[c]
Redick et al. (2016)	attention control	antisaccade (accuracy)	-0.65	.58
		go/no-go (signal-detection measure of performance)	0.48	.77
		go/no-go (standard deviation for go trials)	-0.55	.70
		spatial Stroop (incongruent – congruent, RTs)	-0.28	.92
		cued search (RTs of correct responses)	-0.49	.76
		cued flanker (RTs of incongruent-lure trials)	-0.55	.70
		arrow flanker (incongruent – congruent, RTs)	-0.25	.94
Schweizer & Moosbrugger (2004)	attention	Frankfurt Adaptive Concentration	.91	.17[c]
		Performance Test (RTs)[e]	.92	.16[c]
		Frankfurt Adaptive Concentration	.91	.18[c]
		Performance Test (accuracy)[e]	.90	.19[c]
Schweizer et al. (2005)	common factor[f]	attention: perceptual processing	.92	n.a.

Study	Executive-control label	Measures used to assess executive control	Loadings	Error Variances
		attention: higher mental processing & executive/control processing	.87	n.a.
Shipstead et al. (2014)	attention control	*antisaccade (accuracy)*	*.71*	.49[c]
		arrow flanker (incongruent – neutral, RTs)	-.39	.85[c]
		color Stroop (incongruent – congruent, RTs)	-.25	.94[c]
Unsworth (2015, Exp. 1)	attention control	antisaccade (coefficient of variation for correct RTs)	.50	.75[c]
		arrow flanker (coefficient of variation for correct RTs)	.74	.45[c]
		color Stroop (coefficient of variation for correct RTs)	.73	.47[c]
		psychomotor vigilance (coefficient of variation for correct RTs)	.29	.92[c]
Unsworth (2015, Exp. 2)	attention control	antisaccade (coefficient of variation for correct RTs)	.67	.55[c]
		arrow flanker (coefficient of variation for correct RTs)	.65	.67[c]
		psychomotor vigilance (coefficient of variation for correct RTs)	.57	.58[c]
Unsworth (2015, Exp. 3)	attention control	antisaccade (coefficient of variation for correct RTs)	.32	.90[c]
		arrow flanker (coefficient of variation for correct RTs)	.71	.50[c]
		go/no-go (coefficient of variation for correct RTs)	.34	.88[c]
		color Stroop (coefficient of variation for	.41	.83[c]

Study	Executive-control label	Measures used to assess executive control	Loadings	Error Variances
		correct RTs)		
		psychomotor vigilance (coefficient of variation for correct RTs)	.54	.71[c]
Unsworth & McMillan (2014)	attention control	antisaccade (accuracy)	.54	.71[c]
		arrow flanker (incongruent – congruent, RTs)	-.33	.89[c]
		go/no-go (standard deviation for go trials)	-.45	.80[c]
		go/no-go (accuracy)	.38	.87[c]
		color Stroop (incongruent – congruent, RTs)	-.30	.91[c]
		psychomotor vigilance (average reaction time for the slowest 20% of trials)	-.45	.80[c]
Unsworth, Miller et al. (2009)	response inhibition	*antisaccade (accuracy)*	*.69*	.52
		arrow flanker (incongruent – congruent, RTs)	-.43	.82
Unsworth et al. (2014)	attention control	disengage (difference between the number of the items held in WM for set size 4 and set size 8, accuracy)	-.47	.78[c]
		antisaccade (accuracy)	.62	.62[c]
		48drop (difference between the no-distractor trials and 1/3-distractor trials of the second phase, accuracy)	-.40	.84[c]
Unsworth et al. (2010)	response inhibition	antisaccade (accuracy)	.45	.80
		arrow flanker (incongruent – congruent, RTs)	-.57	.67
Unsworth & Spillers (2010)	attention control	antisaccade (accuracy)	.63	.61[c]
		arrow flanker (incongruent – congruent, RTs)	-.49	.76[c]

Study	Executive-control label	Measures used to assess executive control	Loadings	Error Variances
Unsworth, Spillers et al. (2009)	attention control	color Stroop (incongruent – congruent, RTs)	-.32	.90[c]
		psychomotor vigilance (average reaction time for the slowest 20% of trials)	-.42	.82[c]
		antisaccade (accuracy)	.43	.82[c]
		arrow flanker (incongruent – congruent, RTs)	-.47	.78[c]
		psychomotor vigilance (average reaction time for the slowest 20% of trials)	-.43	.81[c]

Note. As executive control and WMC latent variables were labelled differently across studies, the second and third column refer to the factor names used in the studies and list the different factors if different factors were computed for the executive control and WM constructs. When the factor is dominated by a single measure (i.e., the first or second highest loading was more than 1.5 times as high as the subsequent loading), the measure with the high loading on the factor is presented in italics. When error variances were not reported in the original study, these were computed as follows: 1-(factor loading)^2. n.a. = not available. SD = standard deviation. RT = reaction time.

[a]The backward digit span (i.e., a task in which participants are asked to recall lists of numbers of increasing length in reverse order) and the sentence repetition task (i.e., a task in which participants are asked to repeat the sentence verbatim) are assumed to measure working memory. The letter monitoring task (i.e., a task in which participants are asked to read letters aloud from one side of the computer screen while ignoring letters and numbers on the opposite side and numbers, and then to alternate the side if a cue requires it) is assumed to measure task switching.

[b]The difference score for each participant was expressed as the residual of the incongruent trials regressed on the congruent or neutral trials.

[c]This error variance was reported in the original study.

[d]The factor loadings for the model with executive control and working memory as correlated but distinct factors were not reported; only the factor loadings from the model in which all tasks loaded on a single factor were presented. These values are reported in Table B1.

[e]In the next columns, values in the first row of the cell refers to the model in which gF was measured with the Raven Advanced Progressive Matrices (RAPM); values in the second row of the cell refer to the model in which gF was measured with the "Zahlen-Verbindungs-Test" (ZVT).

[f]The executive-control factor was a second-order factor on which two factors loaded: "attention: perceptual processing" and "attention: higher mental processing & executive/control processing". Four measures loaded on the factor "attention: perceptual processing" (i.e., alertness [loading = .59; error variance = .65], selective/focused attention [loading = .54; error variance = .70], attentional switching [loading = .48; error variance = .77], spatial attention [loading = .55; error variance = .69]). Seven measures loaded on the factor "higher mental processing & executive/control processing" (i.e., supervisory attention [loading = .44; error variance = .81], sustained attention [loading = .50; error variance = .75], attention: assessment tradition I [loading = .65; error variance = .58], attention: assessment tradition II [loading = .78; error variance = .39], attention: assessment tradition III [loading = .60; error variance = .63], inhibition [loading = .58; error variance = .66], planning [loading = .46; error variance = .79], divided attention

[loading = .53; error variance = .72], interference [loading = .70; error variance = .51]).

Appendix C

Block Structure

Table C1

Block order and number of trials per block for each task.

Block order	Trial type / Task	Number of trials per block
Color Stroop, number Stroop, arrow flanker, and letter flanker		
Practice block	incongruent and congruent trials[a]	16
Calibration block 1	neutral trials	28 (incl. 8 warm-up trials)
Experimental block 1	incongruent and congruent trials[a]	66 (incl. 2 warm-up trials)
Calibration block 2	neutral trials	22 (incl. 2 warm-up trials)
Experimental block 2	incongruent and congruent trials[a]	66 (incl. 2 warm-up trials)
Calibration block 3	neutral trials	22 (incl. 2 warm-up trials)
Experimental block 3	incongruent and congruent trials[a]	66 (incl. 2 warm-up trials)
Simon[b]		
Practice block	incongruent and congruent trials[a]	16
Calibration block 1	congruent trials	28 (incl. 8 warm-up trials)
Experimental block 1	incongruent and congruent trials[a]	66 (incl. 2 warm-up trials)
Calibration block 2	congruent trials	22 (incl. 2 warm-up trials)
Experimental block 2	incongruent and congruent trials[a]	66 (incl. 2 warm-up trials)
Calibration block 3	congruent trials	22 (incl. 2 warm-up trials)
Experimental block 3	incongruent and congruent trials[a]	66 (incl. 2 warm-up trials)
Antisaccade		
Calibration block 1	prosaccade trials	50 (incl. 10 warm-up trials)
Practice block 1	antisaccade trials	12
Experimental block 1	antisaccade trials	50 (incl. 2 warm-up trials)
Calibration block 2	prosaccade trials	50 (incl. 10 warm-up trials)
Practice block 2	antisaccade trials	12
Experimental block 2	antisaccade trials	50 (incl. 2 warm-up trials)
Stop-signal		
Practice block 1	go trials	24 trials
Practice block 2	go and stop trials[c]	16 trials
4 experimental blocks	go and stop trials[c]	74 (incl. 2 warm-up trials)
Letter sets		
Practice block	-	2
Experimental block 1	-	15
Experimental block 2	-	15
Locations		

Block order	Trial type / Task	Number of trials per block
Practice block	-	2
Experimental block 1	-	14
Experimental block 2	-	14
RAPM		
Practice block	-	2
Experimental block	-	12
Relationships		
Practice block	-	2
Experimental block 1	-	15
Experimental block 2	-	15
Syllogisms		
Practice block	-	5
Experimental block 1	-	15
Experimental block 2	-	15
Numerical updating		
Practice block	-	3
Experimental block 1	-	13
Experimental block 2	-	12
Spatial updating		
Practice block	-	2
Experimental block 1	-	6
Experimental block 2	-	6
Experimental block 3		6
Numerical complex span		
Practice block	recall task	2
Experimental block 1	recall task	6
Experimental block 2	recall task	6
Spatial complex span		
Practice block	recall task	2
Experimental block 1	recall task	6
Experimental block 2	recall task	6
Time estimation task		
Experimental block 1	trials with timer	30 (incl. 10 warm-up trials)
	trials without timer	25 (incl. 5 warm-up trials)
Experimental block 2	trials with timer	30 (incl. 10 warm-up trials)
	trials without timer	25 (incl. 5 warm-up trials)
Experimental block 3	trials with timer	30 (incl. 10 warm-up trials)
	trials without timer	25 (incl. 5 warm-up trials)

Note. In the analyses, only the (non-warm-up) trials from the experimental blocks were analyzed. RAPM = Raven's Advanced Progressive Matrices.

[a]All trial types occurred with equal frequency.

[b]Because neutral trials are not usually employed in the Simon task (Hommel, 2011), we presented congruent trials in the calibration blocks.

[c]Stop trials were presented in 38% of the trials for the practice block and 33% of the trials for each experimental block.

Appendix D

Experimental Blocks: Error Rates for Incongruent and Congruent trials for the Tasks Assumed to Measure Executive Control

For the color Stroop, number Stroop, arrow flanker, letter flanker and Simon tasks, descriptive statistics are presented separately for incongruent and congruent trials in Table D1. The antisaccade task only included antisaccade trials in the experimental blocks, whose performance is shown in Table 3.

We reasoned that as participants take longer to respond to incongruent trials than to congruent trials in the typical executive-control tasks, this increase in time demand should be translated into higher error rates for incongruent than for congruent trials when a deadline approach was used to limit the response time window. As shown in Table D1, this was the case for all tasks. Performance was close to the chance level for the incongruent trials of the letter flanker and Simon tasks. However, removing these tasks from structural equation modeling did not change the pattern of results.

Table D1

Descriptive Statistics for Performance in Incongruent and Congruent Trials.

Task / Trial type	*M*	*SD*	Min.	Max.	Skew	Kurtosis	Reliability
Color Stroop							
Incongruent	25.85	10.22	8.33	80.21	-1.40	4.35	.85
Congruent	17.42	8.01	1.04	70.83	-2.56	12.70	.77
Number Stroop							
Incongruent	34.84	10.57	6.25	65.63	-0.11	-0.34	.81
Congruent	13.86	6.77	2.08	44.79	-1.15	2.25	.78
Arrow flanker							
Incongruent	52.03	15.29	17.71	92.71	-0.17	-0.55	.92
Congruent	12.24	8.28	0.00	60.42	-2.51	9.97	.89
Letter flanker							
Incongruent	26.36	8.63	7.29	52.08	-0.65	0.34	.72
Congruent	17.99	6.85	1.04	46.88	-0.76	1.26	.70
Simon							
Incongruent	56.05	14.56	16.67	93.75	0.19	-0.38	.93
Congruent	35.39	10.31	11.46	66.67	-0.44	-0.03	.82

Note. Reliabilities were calculated by adjusting split-half correlations with the Spearman–Brown prophecy formula. Split-half correlations were computed between odd and even items. Min. = minimum; Max. = maximum.

Appendix E

Performance in the Calibration Blocks

We analyzed performance in the calibration blocks to verify that the calibration procedure worked. In these blocks, two measures are of interest: the accuracy performance and the individually calibrated response deadlines.

Accuracy Performance

Successful calibration should be reflected in two criteria. First, mean accuracy should be close to the criterion. Ideally, this should be true for each individual participant, reducing the variance to zero, but this is unrealistic because the calibration does not remove measurement noise. Successful calibration does remove systematic (i.e., "true") variance between individuals, so that the remaining variance should be predominantly measurement noise. Therefore, our second criterion is that the reliability of performance in the calibration blocks should be low.

The descriptive statistics for the performance in calibration blocks are presented in Table E1. Individual performance in the different blocks are presented for each task separately in Figure E1. As shown in both table and figure, the accuracy rates were approximately 80.76% for the tasks with 75% as cut-off (i.e., for the color Stroop, number Stroop, arrow flanker, letter flanker and Simon tasks). The accuracy rates were 84.36% for the antisaccade task in which the cut-off was 80%. Although most participants' accuracies were reasonably close to the criterion, there was still substantial variance left, justifying our decision to use the difference scores between incongruent and congruent trials for the color Stroop, number Stroop, arrow flanker, letter flanker and Simon tasks, as well as the difference scores between prosaccades and antisaccades trials for the antisaccade task.

The data were not normally distributed (see Table E1, upper part), and this may affect the reliability estimates computed with Cronbach's alpha (Sheng & Sheng, 2012). We checked for multivariate normality using Mardia's (1970) kurtosis index and we removed four multivariate outliers (i.e., cases with significant Mahalanobis's d2 values). This resulted in approximately normally distributed data (see Table E1, lower part). With outliers removed, the reliability estimates were low, indicating that there is little if any systematic variance left in these measures. This result supports that the calibration achieved the goal to reduce true individual difference variance. Please note that removing these multivariate outliers when computing structure equation modeling on the data from the experimental blocks did not improve the results.

Response Deadlines

For the sake of completeness and replicability, we also report the analyses on response deadlines. The descriptive statistics are presented in Table E2. As shown in the upper part of this table, the data were not normally distributed. We checked for multivariate normality using Mardia's (1970) kurtosis index and we removed seven multivariate outliers (i.e., cases with significant Mahalanobis's d2 values). This resulted in approximately normally distributed data (see Table E2, lower part). In all cases, reliabilities were good.

In order to verify whether individuals with higher WMC or gF were better calibrated or more responsive to the response deadlines than were individuals with lower WMC or gF, we computed the correlations between the response deadlines and the measures of WMC and gF. The correlations are presented in Table E3. Table E4 presents the Bayes Factor in favor of the alternative hypothesis (BF_{10}, i.e., in favor of a correlation) and the Bayes Factor in favor of the null hypothesis (BF_{01}, i.e., in favor of the absence of the correlation). Most correlations were low and non-significant, and most BF_{01} show positive to strong evidence for the absence of correlations. Together, these findings ruled out that individual differences in WMC and/or gF were related to how participants handled the response deadline. Please note that removing these multivariate outliers when computing structure equation modeling on the data from the experimental blocks did not improve the results.

Table E1

Descriptive Statistics for the Accuracy Rates in the Calibration Blocks.

Trimming / Task	Sample Size	*M*	*SD*	Min.	Max.	Skew	Kurtosis	Reliability
Sample as in the main analysis								
Color Stroop	181	80.10	5.72	31.67	90.00	-4.54	32.63	.54
Number Stroop	181	80.48	3.83	61.67	91.67	-1.34	4.85	.17
Arrow flanker	181	80.45	4.23	43.33	88.33	-4.75	36.06	.35
Letter flanker	181	80.75	4.70	48.33	91.67	-2.37	12.87	.40
Simon	181	82.00	2.92	68.33	90.00	-0.94	2.95	.22
Antisaccade	181	84.36	5.03	45.00	93.75	-3.80	24.05	.45
Without multivariate outliers								
Color Stroop	177	80.58	3.67	63.33	90.00	-0.71	2.00	-.03
Number Stroop	177	80.64	3.39	65.00	91.67	-0.62	2.36	.07
Arrow flanker	177	80.79	2.67	70.00	88.33	-0.50	1.44	-.02
Letter flanker	177	81.11	3.68	66.67	91.67	-0.49	1.56	.15
Simon	177	82.10	2.76	70.00	90.00	-0.56	1.32	.21
Antisaccade	177	84.75	3.57	67.50	93.75	-1.21	4.92	.14

Note. Scores were computed as accuracy rates (in %) in neutral trials for the color Stroop, number Stroop, arrow flanker, letter flanker, in congruent trials for the Simon task, and in prosaccade trials for the antisaccade task. Because of the tracking procedure on the response deadline in the calibration blocks, we computed the Cronbach's alpha across the three calibration blocks for the color Stroop, number Stroop, arrow flanker, letter flanker and Simon tasks and across the two calibration blocks for the antisaccade tasks. Min. = minimum; Max. = maximum.

Table E2

Descriptive Statistics for the Response Deadlines.

Trimming / Task	Sample Size	*M*	*SD*	Min.	Max.	Skew	Kurtosis	Reliability
Sample as in the main analysis								
Color Stroop	181	740	128	537	1839	3.68	28.47	.91
Number Stroop	181	628	94	492	1161	1.64	4.97	.89
Arrow flanker	181	440	70	350	1093	4.96	40.29	.79
Letter flanker	181	633	124	472	1443	3.00	13.44	.89
Simon	181	386	52	298	734	2.51	12.02	.79
Antisaccade	181	157	107	31	737	2.64	8.59	.79
Without multivariate outliers								
Color Stroop	174	730	93	537	1023	0.39	0.10	.86
Number Stroop	174	622	83	492	891	0.97	0.72	.86
Arrow flanker	174	433	46	350	617	1.16	1.98	.76
Letter flanker	174	620	88	472	999	1.41	2.71	.86
Simon	174	382	41	298	552	1.05	1.83	.72
Antisaccade	174	146	80	31	500	1.87	4.33	.67

Note. Median response deadlines were averaged across the calibration blocks. Because of the tracking procedure on the response deadline in the calibration blocks, we computed the Cronbach's alpha across the three calibration blocks for the color Stroop, number Stroop, arrow flanker, letter flanker and Simon tasks and across the two calibration blocks for the antisaccade tasks. Min. = minimum; Max. = maximum.

Table E3

Response Deadlines: Pearson Correlation Coefficients.

Task	Letter sets	Loca -tions	RAPM	Relation ships	Syllo -gisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.s.
Color	.21*	*.04*	*.08*	*.10*	*-.02*	**.24***	**.26***	*.03*	*.12*
Stroop	[.07, .35]	[-.11, .20]	[-.07, .22]	[-.04, .25]	[-.16, .13]	[.09, .39]	[.14, .39]	[-.13, .18]	[-.03, .27]
Number	*.07*	*-.001*	.18*	*-.002*	*-.10*	**.33***	*-.003*	*.07*	**.29***
Stroop	[-.06, .20]	[-.13, .13]	[.04, .31]	[-.13, .13]	[-.25, .04]	[.21, .44]	[-.15, .14]	[-.09, .24]	[.16, .42]
Arrow	*.12*	*-.09*	.16*	*.11*	*-.04*	*.03*	*-.01*	*-.05*	*.11*
flanker	[-.06, .29]	[-.24, .06]	[.01, .30]	[-.06, .27]	[-.20, .11]	[-.13, .19]	[-.16, .13]	[-.20, .10]	[-.04, .25]
Letter	.19*	*.04*	.15*	*.07*	*-.06*	*.13*	*.06*	*-.001*	.20*
flanker	[.04, .35]	[-.11, .20]	[.01, .30]	[-.10, .24]	[-.22, .10]	[-.01, .27]	[-.08, .21]	[-.16, .16]	[.07, .33]
Simon	.20*	*.10*	*.08*	.14	*.04*	.17*	*-.02*	*.07*	.17*
	[.03, .37]	[-.04, .25]	[-.08, .24]	[-.06, .36]	[-.10, .17]	[.03, .30]	[-.16, .12]	[-.09, .23]	[.03, .31]
Anti	*.10*	**.24***	.14	*.07*	*.11*	.20*	**.22***	*.10*	**.25***
-saccade	[-.02, .23]	[.12, .36]	[.02, .27]	[-.07, .20]	[-.05, .27]	[.07, .33]	[.09, .33]	[-.06, .27]	[.12, .38]

Note. Ninety-five % bootstrapped confidence intervals (10000 random samples) are presented in brackets. Correlations for which the Bayes factor suggested positive to strong evidence for the alternative hypothesis (BF_{10}) are presented in bold; correlations for which the Bayes factor suggested positive to strong evidence for the null hypothesis (BF_{01}) are presented in italics. Bayes factors for each correlation are presented in Table E4. Num. = Numerical; upd. = updating; c.s. = complex span.

* $p < .05$.

Table E4

Response Deadlines: Bayes Factors in favor of the Alternative Hypothesis (BF_{10}) and in favor of the Null hypothesis (BF_{01}) for the Pearson Correlation Coefficients.

Task	BF	Letter sets	Loca -tions	RAPM	Relation ships	Syllo -gisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.s.
Color	BF_{10}	**2.75**	**0.07**	**0.10**	**0.15**	**0.06**	**10.52**	**26.00**	**0.06**	**0.19**
Stroop	BF_{01}	*0.36*	*14.00*	*10.12*	*6.49*	*16.29*	*0.10*	*0.04*	*15.73*	*5.15*
Number	BF_{10}	**0.09**	**0.06**	**0.87**	**0.06**	**0.15**	**908.87**	**0.06**	**0.10**	**128.02**
Stroop	BF_{01}	*10.92*	*16.62*	*1.15*	*16.62*	*6.79*	*1.10e-03*	*16.61*	*10.38*	*0.01*
Arrow	BF_{10}	**0.19**	**0.12**	**0.51**	**0.16**	**0.07**	**0.06**	**0.06**	**0.07**	**0.16**
flanker	BF_{01}	*5.24*	*8.53*	*1.96*	*6.30*	*14.03*	*15.61*	*16.34*	*13.52*	*6.31*
Letter	BF_{10}	**1.48**	**0.07**	**0.45**	**0.09**	**0.08**	**0.25**	**0.08**	**0.06**	**1.94**
flanker	BF_{01}	*0.67*	*13.99*	*2.21*	*10.71*	*12.41*	*3.97*	*11.83*	*16.63*	*0.52*
Simon	BF_{10}	**1.91**	**0.15**	**0.10**	**0.34**	**0.07**	**0.72**	**0.06**	**0.09**	**0.70**
	BF_{01}	*0.52*	*6.60*	*10.07*	*2.91*	*14.93*	*1.39*	*16.13*	*10.73*	*1.44*
Anti	BF_{10}	**0.15**	**8.15**	**0.37**	**0.09**	**0.18**	**1.88**	**3.61**	**0.15**	**16.82**
-saccade	BF_{01}	*6.60*	*0.12*	*2.70*	*11.54*	*5.67*	*0.53*	*0.28*	*6.55*	*0.06*

Note. For the sake of clarity, BF_{10} are presented in bold, whereas BF_{01} are presented in italics. BF = Bayes Factor; num. = numerical; upd. = updating; c.s. = complex span.

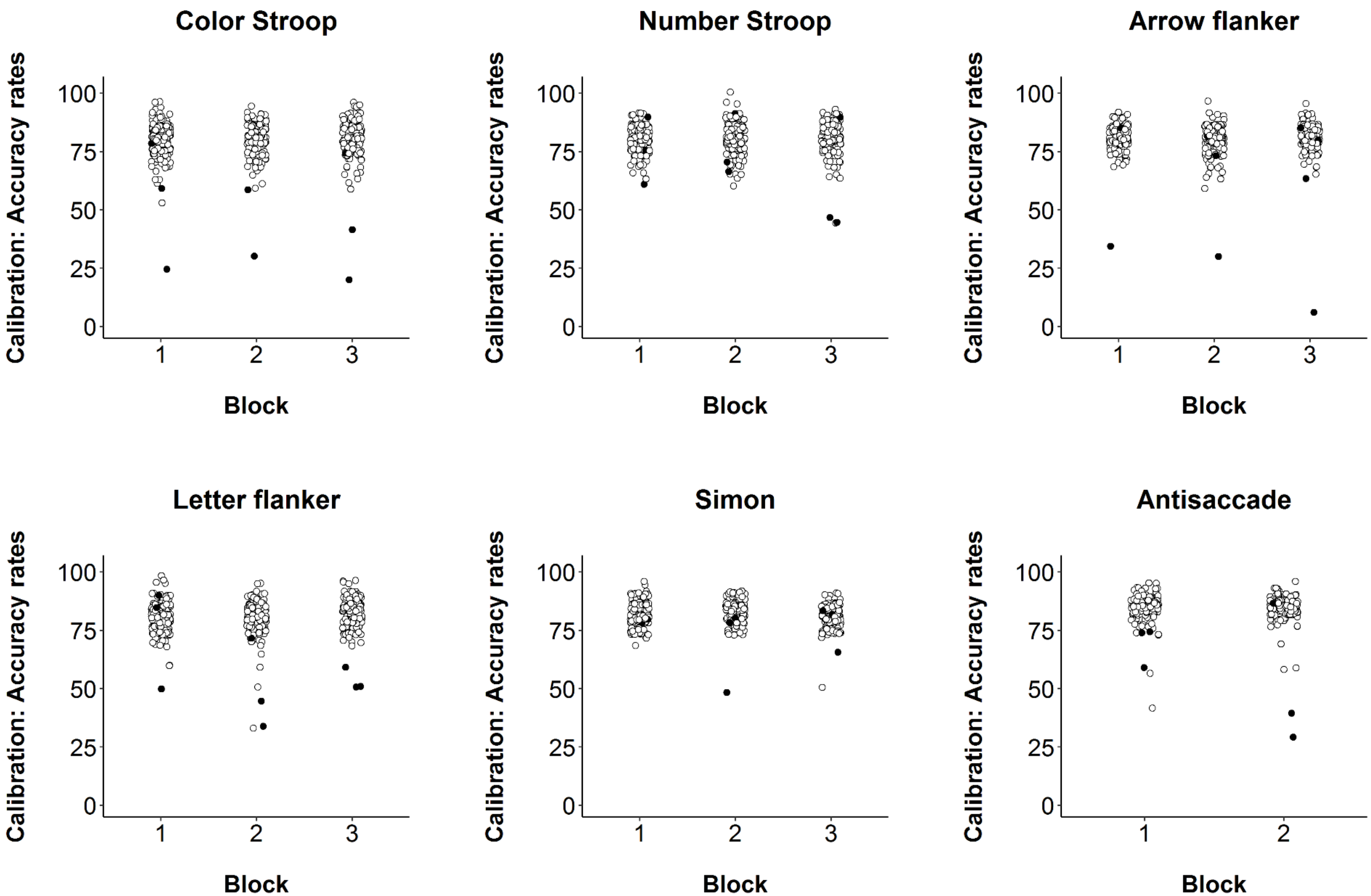

Figure E1. Individual accuracy performance in the three calibration blocks in each task. Black dots represent multivariate outliers.

Appendix F

Errors of Omission and Commission in the Executive-Control Tasks

Following the rationale that any failure to produce a correct response within the allotted time given by the response deadline – whether by giving a wrong response or by missing the deadline – reflects a lack of executive-control ability, we treated errors of commission (i.e., giving a wrong response) and errors of omission (i.e., failing to respond before the deadline) as equivalent. In the present study, this concerned all tasks assumed to measure executive control, except the stop-signal task which yields only errors of omissions. One may argue, however, that individual differences may occur in how the response deadline is handled. For example, it is possible that participants with higher WMC/gF are better at meeting the response deadline than participants with lower WMC/gF.

To ensure that both error types did no correlate with any variable of interest, we computed error rates separately for errors of omission and errors of commission. The descriptive statistics for both error types are presented in Table F1. As the data were not normally distributed (see Table F1, upper part), we checked for multivariate normality using Mardia's (1970) kurtosis index, and we removed seven multivariate outliers (i.e., cases with significant Mahalanobis's d2 values). This yielded approximately normally distributed data (see Table F1, lower part). In most cases, reliabilities were good to acceptable. Only the reliability for the omission errors of the Simon task in the calibration block and the reliability for the commission errors of the antisaccade tasks in the calibration block were unacceptably low (.42 and .28, respectively).

The correlations between the two error rates calculated with both types of errors and the measures of WMC and gF are presented in Table F2. Table F3 presents the Bayes Factor in favor of the alternative hypothesis (BF_{10}, i.e., in favor of a correlation) and the Bayes Factor in favor of the null hypothesis (BF_{01}, i.e., in favor of the absence of the correlation). Most correlations were low and non-significant, and most BF_{01} show positive to strong evidence for the absence of correlations. Together, these findings ruled out that individual differences in WMC and/or gF were related to how participants handled the response deadline.

Table F1

Descriptive Statistics for Errors of Omission and of Commission.

Trimming / Type of block	Type of error	Task	Sample Size	*M*	*SD*	Min.	Max.	Skew	Kurtosis	Reliability
Sample as in the main analysis										
Experimental blocks	Omission	Color Stroop	181	12.29	0.06	0.45	34.88	0.57	0.74	.87
		Number Stroop	181	13.41	0.07	1.82	41.47	0.80	0.99	.91
		Arrow flanker	181	20.66	0.09	0.81	43.49	0.06	-0.68	.92
		Letter flanker	181	10.53	0.06	0.00	31.17	0.62	0.34	.89
		Simon	181	32.13	0.08	11.99	55.74	0.11	-0.27	.88
		Antisaccade	181	4.73	0.07	0.00	40.49	2.76	8.08	.94
	Commission	Color Stroop	181	11.14	0.08	0.49	61.19	3.32	18.46	.93
		Number Stroop	181	12.44	0.06	1.87	34.93	0.85	1.22	.85
		Arrow flanker	181	13.77	0.07	1.01	58.39	2.08	10.38	.90
		Letter flanker	181	12.76	0.07	1.08	47.50	1.54	5.04	.89
		Simon	181	12.00	0.07	1.63	33.81	0.63	-0.01	.91
		Antisaccade	181	24.85	0.10	4.45	61.13	0.32	0.19	.91
Calibration blocks	Omission	Color Stroop	181	10.98	0.05	0.00	29.63	0.09	0.15	.57
		Number Stroop	181	11.06	0.05	0.00	23.73	-0.18	-0.50	.60
		Arrow flanker	181	12.73	0.04	0.00	21.05	-0.44	-0.12	.49
		Letter flanker	181	8.80	0.05	0.00	20.69	0.03	-0.70	.58
		Simon	181	16.46	0.03	5.26	27.59	-0.14	0.47	.37
		Antisaccade	181	1.88	0.03	0.00	14.10	1.92	3.67	.48
	Commission	Color Stroop	181	10.77	0.07	0.00	63.46	2.83	15.73	.69
		Number Stroop	181	10.08	0.06	0.00	36.21	0.95	1.83	.57
		Arrow flanker	181	8.41	0.07	0.00	51.85	2.33	10.78	.58
		Letter flanker	181	11.91	0.07	0.00	50.85	1.57	5.58	.72
		Simon	181	2.05	0.03	0.00	19.61	2.27	6.14	.60
		Antisaccade	181	14.18	0.05	2.90	53.85	3.05	18.32	.49
Without multivariate outliers										

Trimming / Type of block	Type of error	Task	Sample Size	*M*	*SD*	Min.	Max.	Skew	Kurtosis	Reliability
Experimental blocks	Omission	Color Stroop	174	11.98	0.05	0.45	29.64	0.33	-0.04	.85
		Number Stroop	174	13.54	0.07	1.82	41.47	0.81	1.01	.90
		Arrow flanker	174	20.37	0.09	0.81	43.49	0.09	-0.70	.92
		Letter flanker	174	10.56	0.06	0.00	31.17	0.63	0.42	.89
		Simon	174	31.97	0.08	11.99	55.74	0.11	-0.30	.88
		Antisaccade	174	4.46	0.07	0.00	39.50	2.81	8.45	.93
	Commission	Color Stroop	174	10.52	0.05	0.49	32.67	0.81	0.89	.87
		Number Stroop	174	12.20	0.06	1.87	34.93	0.83	1.36	.84
		Arrow flanker	174	13.22	0.05	1.01	29.25	0.07	-0.47	.84
		Letter flanker	174	12.28	0.06	1.08	35.25	0.65	0.77	.85
		Simon	174	11.68	0.06	1.63	28.06	0.40	-0.71	.90
		Antisaccade	174	24.29	0.09	4.45	44.44	0.03	-0.65	.90
Calibration blocks	Omission	Color Stroop	174	10.77	0.05	0.00	23.33	-0.10	-0.47	.53
		Number Stroop	174	11.20	0.05	0.00	23.73	-0.20	-0.41	.60
		Arrow flanker	174	12.68	0.04	0.00	21.05	-0.44	-0.08	.52
		Letter flanker	174	8.90	0.05	0.00	20.69	0.01	-0.68	.57
		Simon	174	16.45	0.03	5.26	27.59	-0.13	0.46	.42
		Antisaccade	174	1.81	0.03	0.00	14.10	1.98	4.00	.58
	Commission	Color Stroop	174	10.18	0.06	0.00	32.14	0.64	0.64	.55
		Number Stroop	174	9.73	0.06	0.00	36.21	0.89	1.82	.59
		Arrow flanker	174	7.95	0.05	0.00	25.00	0.76	0.16	.55
		Letter flanker	174	11.45	0.06	0.00	31.67	0.67	0.64	.66
		Simon	174	1.94	0.03	0.00	14.55	1.98	3.85	.56
		Antisaccade	174	13.76	0.04	2.90	31.65	0.67	2.63	.28

Note. Scores for errors of omission and of commission were computed as error rates (in %) averaged across incongruent and congruent trials in the experimental blocks and error rates of neutral trials in the calibration blocks for the color Stroop, number Stroop, arrow flanker, and letter flanker task. For the Simon task, scores were computed as error rates (in %) averaged across incongruent and congruent trials in the experimental blocks and error rates of congruent trials in the calibration blocks. For the antisaccade task, scores were computed as error rates (in %) of antisaccade trials in the experimental blocks and error rates of prosaccade trials in the calibration blocks. Reliabilities for the experimental blocks were calculated by adjusting split-half correlations with the Spearman–Brown prophecy formula. Split-half correlations were computed between odd and even items. Reliabilities for the calibration blocks were computed with Cronbach's alpha across the calibration blocks. Min. = minimum; Max. = maximum.

Table F2

Pearson Correlation Coefficients.

Block / Error	Task	Letter sets	Loca -tions	RAPM	Relation ships	Syllo -gisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.s.
Experimental										
Omis	Color	-.02	-.01	-.01	.03	.12	-.05	-.01	.04	-.02
-sion	Stroop	[-.15, .12]	[-.15, .13]	[-.14, .12]	[-.12, .17]	[-.02, .26]	[-.19, .10]	[-.16, .15]	[-.11, .20]	[-.16, .11]
	Number	.21*	.13	.06	.17*	.14	-.04	.12	.003	-.02
	Stroop	[.02, .41]	[-.02, .28]	[-.13, .25]	[-.002, .34]	[.01, .28]	[-.20, .12]	[-.04, .29]	[-.14, .14]	[-.16, .12]
	Arrow	.05	.16*	-.02	.003	.05	**.21***	.07	.17*	.16*
	flanker	[-.10, .19]	[.01, .32]	[-.17, .13]	[-.15, .16]	[-.12, .21]	[.07, .36]	[-.07, .21]	[.01, .32]	[.03, .29]
	Letter	-.10	-.04	-.10	-.07	.09	-.07	-.02	.002	-.11
	flanker	[-.24, .05]	[-.17, .10]	[-.25, .04]	[-.21, .07]	[-.06, .24]	[-.21, .07]	[-.18, .13]	[-.15, .16]	[-.24, .03]
	Simon	.004	-.04	.04	.08	-.05	-.05	.15*	.02	.01
		[-.15, .16]	[-.20, .12]	[-.11, .19]	[-.08, .24]	[-.18, .09]	[-.21, .11]	[.000, .31]	[-.12, .16]	[-.15, .16]
	Anti	-.04	-.11	.06	-.11	.02	.01	-.10	.07	-.08
	-saccade	[-.18, .09]	[-.23, -.002]	[-.08, .20]	[-.24, .01]	[-.15, .19]	[-.13, .15]	[-.21, .003]	[-.10, .23]	[-.26, .09]
Commis	Color	-.01	-.005	.06	.05	-.12	.10	-.04	.11	.18*
-sion	Stroop	[-.16, .14]	[-.15, .14]	[-.09, .21]	[-.09, .20]	[-.27, .03]	[-.04, .24]	[-.19, .10]	[-.05, .27]	[.03, .33]
	Number	.02	-.03	.13	.02	-.05	.09	-.05	-.004	.13
	Stroop	[-.12, .17]	[-.18, .11]	[-.01, .28]	[-.11, .14]	[-.21, .12]	[-.05, .24]	[-.18, .09]	[-.15, .14]	[-.01, .28]
	Arrow	.06	.003	-.01	.004	-.05	-.05	-.06	-.03	.01
	flanker	[-.11, .23]	[-.14, .15]	[-.17, .15]	[-.15, .16]	[-.21, .10]	[-.20, .09]	[-.22, .10]	[-.18, .12]	[-.15, .16]
	Letter	.05	.03	.10	.09	-.10	.11	.003	.03	.19*
	flanker	[-.11, .20]	[-.09, .16]	[-.06, .26]	[-.05, .23]	[-.26, .07]	[-.03, .24]	[-.15, .15]	[-.11, .18]	[.05, .34]
	Simon	-.04	-.10	.03	-.01	-.12	.01	.003	-.04	.05
		[-.19, .11]	[-.24, .05]	[-.13, .18]	[-.16, .13]	[-.27, .03]	[-.14, .16]	[-.15, .15]	[-.20, .11]	[-.11, .20]
	Anti	.09	.000	.06	.10	-.02	.04	-.01	.03	.04
	-saccade	[-.05, .22]	[-.15, .15]	[-.08, .20]	[-.05, .26]	[-.16, .13]	[-.12, .19]	[-.15, .14]	[-.13, .19]	[-.11, .19]

Block / Error	Task	Letter sets	Loca -tions	RAPM	Relation ships	Syllo -gisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.s.
Calibration										
Omis -sion	Color Stroop	*.03*	*.08*	*.04*	*-.01*	.15	*-.04*	*.01*	*-.01*	*-.13*
		[-.11, .17]	[-.06, .22]	[-.09, .18]	[-.15, .13]	[.01, .28]	[-.19, .11]	[-.15, .16]	[-.17, .15]	[-.28, .02]
	Number Stroop	.16*	*.09*	*-.01*	*.10*	*.04*	*-.12*	*.11*	*-.01*	-.14
		[.004, .32]	[-.05, .23]	[-.17, .15]	[-.05, .25]	[-.11, .20]	[-.26, .04]	[-.03, .26]	[-.15, .12]	[-.29, -.003]
	Arrow flanker	*-.04*	.14	*-.07*	*-.004*	*.11*	.15	*.02*	*.06*	*.06*
		[-.22, .14]	[-.02, .31]	[-.23, .10]	[-.18, .17]	[-.07, .29]	[-.01, .31]	[-.14, .17]	[-.09, .21]	[-.09, .21]
	Letter flanker	*-.11*	*.01*	*-.07*	*.02*	*.08*	*-.13*	*-.02*	*-.10*	-.18*
		[-.25, .02]	[-.13, .14]	[-.22, .07]	[-.12, .17]	[-.08, .24]	[-.27, .01]	[-.16, .13]	[-.25, .06]	[-.33, -.03]
	Simon	*.07*	*.04*	*-.06*	.15*	.15*	*-.06*	*.07*	*.01*	*-.01*
		[-.10, .25]	[-.10, .18]	[-.22, .11]	[-.03, .34]	[.02, .28]	[-.21, .09]	[-.06, .21]	[-.13, .14]	[-.16, .13]
	Anti -saccade	*-.07*	*-.03*	*.03*	*-.08*	*.11*	*.07*	*-.06*	*.09*	*-.03*
		[-.20, .06]	[-.17, .10]	[-.10, .17]	[-.21, .06]	[-.03, .25]	[-.08, .22]	[-.21, .08]	[-.07, .25]	[-.19, .14]
Commis -sion	Color Stroop	*-.03*	*.04*	*.01*	*.07*	-.19*	*.09*	*-.03*	*.09*	*.12*
		[-.18, .12]	[-.12, .19]	[-.13, .15]	[-.06, .20]	[-.33, -.04]	[-.05, .23]	[-.18, .11]	[-.07, .26]	[-.03, .27]
	Number Stroop	*-.02*	*-.002*	*.05*	*-.04*	*.04*	*.13*	*-.05*	*-.03*	.14
		[-.18, .13]	[-.14, .14]	[-.11, .21]	[-.18, .10]	[-.11, .19]	[-.02, .27]	[-.19, .09]	[-.16, .11]	[-.001, .28]
	Arrow flanker	*.05*	*-.12*	*-.001*	*-.03*	*-.10*	**-.22***	*-.09*	-.15	*-.13*
		[-.10, .20]	[-.27, .03]	[-.16, .16]	[-.18, .11]	[-.27, .07]	[-.37, -.08]	[-.23, .05]	[-.29, .002]	[-.27, .02]
	Letter flanker	*.14*	*.07*	.15	*.08*	*-.08*	*.12*	*.01*	*.03*	.21*
		[-.01, .29]	[-.06, .20]	[.002, .29]	[-.07, .23]	[-.23, .07]	[-.02, .25]	[-.13, .15]	[-.11, .17]	[.07, .35]
	Simon	*-.05*	-.14	*-.04*	*-.10*	-.18*	*-.09*	-.16*	*-.10*	*-.06*
		[-.21, .10]	[-.30, .02]	[-.20, .11]	[-.26, .06]	[-.33, -.02]	[-.24, .06]	[-.30, -.02]	[-.26, .05]	[-.21, .09]
	Anti -saccade	.17*	*.07*	*.10*	*.01*	*-.08*	*.002*	*.06*	*-.01*	*.12*
		[.03, .32]	[-.07, .21]	[-.04, .25]	[-.12, .15]	[-.21, .05]	[-.14, .14]	[-.07, .18]	[-.16, .14]	[-.03, .27]

Note. Ninety-five % bootstrapped confidence intervals (10000 random samples) are presented in brackets. Correlations for which the Bayes factor suggested positive to strong evidence for the alternative hypothesis (BF_{10}) are presented in bold; correlations for which the Bayes factor suggested positive to strong evidence for the null hypothesis (BF_{01}) are presented in italics. Bayes factors for each correlation are presented in Table F3. Num. = Numerical; upd. = updating; c.s. = complex span.

* $p < .05$.

Table F3

Bayes Factors in favor of the Alternative Hypothesis (BF_{10}) and in favor of the Null hypothesis (BF_{01}) for the Pearson Correlation Coefficients.

Block / Error	Task	BF	Letter sets	Loca -tions	RAPM	Relation ships	Syllo -gisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.s.
Experimental											
Omis -sion	Color Stroop	BF_{10}	**0.06**	**0.06**	**0.06**	**0.06**	**0.20**	**0.07**	**0.06**	**0.07**	**0.06**
		BF_{01}	*16.28*	*16.54*	*16.38*	*15.60*	*5.06*	*13.71*	*16.55*	*14.20*	*15.78*
	Number Stroop	BF_{10}	**2.82**	**0.28**	**0.08**	**0.68**	**0.36**	**0.07**	**0.22**	**0.06**	**0.06**
		BF_{01}	*0.36*	*3.64*	*12.72*	*1.46*	*2.76*	*14.11*	*4.60*	*16.61*	*15.85*
	Arrow flanker	BF_{10}	**0.07**	**0.63**	**0.06**	**0.06**	**0.07**	**3.24**	**0.09**	**0.64**	**0.58**
		BF_{01}	*13.89*	*1.58*	*16.16*	*16.61*	*13.97*	*0.31*	*10.72*	*1.57*	*1.73*
	Letter flanker	BF_{10}	**0.13**	**0.07**	**0.15**	**0.09**	**0.12**	**0.09**	**0.06**	**0.06**	**0.16**
		BF_{01}	*7.43*	*14.89*	*6.71*	*11.32*	*8.68*	*10.97*	*16.00*	*16.62*	*6.10*
	Simon	BF_{10}	**0.06**	**0.07**	**0.07**	**0.10**	**0.07**	**0.07**	**0.47**	**0.06**	**0.06**
		BF_{01}	*16.60*	*14.72*	*14.56*	*9.77*	*13.96*	*13.71*	*2.11*	*15.98*	*16.59*
	Anti -saccade	BF_{10}	**0.07**	**0.19**	**0.08**	**0.18**	**0.06**	**0.06**	**0.14**	**0.09**	**0.10**
		BF_{01}	*14.08*	*5.40*	*12.25*	*5.45*	*15.86*	*16.48*	*6.95*	*10.97*	*9.56*
Commis -sion	Color Stroop	BF_{10}	**0.06**	**0.06**	**0.08**	**0.08**	**0.22**	**0.14**	**0.07**	**0.16**	**0.99**
		BF_{01}	*16.50*	*16.60*	*12.13*	*13.07*	*4.58*	*7.17*	*14.20*	*6.36*	*1.01*
	Number Stroop	BF_{10}	**0.06**	**0.07**	**0.28**	**0.06**	**0.07**	**0.13**	**0.07**	**0.06**	**0.29**
		BF_{01}	*15.91*	*15.03*	*3.54*	*16.04*	*13.49*	*7.80*	*13.38*	*16.61*	*3.50*
	Arrow flanker	BF_{10}	**0.08**	**0.06**	**0.06**	**0.06**	**0.08**	**0.08**	**0.08**	**0.06**	**0.06**
		BF_{01}	*12.00*	*16.62*	*16.39*	*16.61*	*13.12*	*13.18*	*12.34*	*15.62*	*16.59*
	Letter flanker	BF_{10}	**0.07**	**0.07**	**0.14**	**0.12**	**0.14**	**0.16**	**0.06**	**0.07**	**1.47**
		BF_{01}	*13.86*	*15.29*	*7.16*	*8.21*	*7.35*	*6.29*	*16.61*	*15.19*	*0.68*
	Simon	BF_{10}	**0.07**	**0.13**	**0.06**	**0.06**	**0.20**	**0.06**	**0.06**	**0.07**	**0.07**
		BF_{01}	*14.90*	*7.57*	*15.64*	*16.47*	*5.12*	*16.53*	*16.61*	*14.31*	*13.38*

Block / Error	Task	BF	Letter sets	Loca -tions	RAPM	Relation ships	Syllo -gisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.s.
	Anti	BF_{10}	**0.11**	**0.06**	**0.08**	**0.15**	**0.06**	**0.07**	**0.06**	**0.06**	**0.07**
	-saccade	BF_{01}	*8.94*	*16.63*	*12.68*	*6.89*	*16.27*	*14.76*	*16.59*	*15.54*	*14.70*
Calibration											
Omis	Color	BF_{10}	**0.07**	**0.11**	**0.07**	**0.06**	**0.39**	**0.07**	**0.06**	**0.06**	**0.28**
-sion	Stroop	BF_{01}	*15.35*	*9.32*	*14.19*	*16.52*	*2.58*	*14.24*	*16.57*	*16.39*	*3.56*
	Number	BF_{10}	**0.51**	**0.13**	**0.06**	**0.14**	**0.07**	**0.19**	**0.18**	**0.06**	**0.36**
	Stroop	BF_{01}	*1.94*	*7.78*	*16.45*	*7.10*	*14.00*	*5.32*	*5.62*	*16.39*	*2.74*
	Arrow	BF_{10}	**0.07**	**0.34**	**0.09**	**0.06**	**0.16**	**0.40**	**0.06**	**0.08**	**0.08**
	flanker	BF_{01}	*14.55*	*2.94*	*11.15*	*16.61*	*6.10*	*2.47*	*16.14*	*12.46*	*12.40*
	Letter	BF_{10}	**0.17**	**0.06**	**0.10**	**0.06**	**0.10**	**0.27**	**0.06**	**0.13**	**0.94**
	flanker	BF_{01}	*5.85*	*16.52*	*10.38*	*15.94*	*10.14*	*3.67*	*16.28*	*7.63*	*1.07*
	Simon	BF_{10}	**0.10**	**0.07**	**0.08**	**0.44**	**0.45**	**0.08**	**0.10**	**0.06**	**0.06**
		BF_{01}	*10.42*	*14.58*	*12.42*	*2.28*	*2.24*	*12.80*	*10.32*	*16.55*	*16.34*
	Anti	BF_{10}	**0.09**	**0.07**	**0.07**	**0.10**	**0.17**	**0.09**	**0.08**	**0.12**	**0.06**
	-saccade	BF_{01}	*10.97*	*15.19*	*15.24*	*9.58*	*5.96*	*10.65*	*12.02*	*8.42*	*15.74*
Commis	Color	BF_{10}	**0.07**	**0.07**	**0.06**	**0.09**	**1.23**	**0.12**	**0.07**	**0.12**	**0.21**
-sion	Stroop	BF_{01}	*15.33*	*14.72*	*16.47*	*11.08*	*0.81*	*8.50*	*15.21*	*8.38*	*4.75*
	Number	BF_{10}	**0.06**	**0.06**	**0.07**	**0.07**	**0.07**	**0.24**	**0.07**	**0.06**	**0.33**
	Stroop	BF_{01}	*15.89*	*16.62*	*13.91*	*14.24*	*14.45*	*4.14*	*13.83*	*15.70*	*3.00*
	Arrow	BF_{10}	**0.07**	**0.22**	**0.06**	**0.07**	**0.14**	**4.82**	**0.11**	**0.37**	**0.24**
	flanker	BF_{01}	*13.47*	*4.63*	*16.62*	*15.11*	*7.06*	*0.21*	*8.71*	*2.67*	*4.26*
	Letter	BF_{10}	**0.29**	**0.09**	**0.38**	**0.10**	**0.10**	**0.19**	**0.06**	**0.06**	**2.77**
	flanker	BF_{01}	*3.44*	*10.96*	*2.61*	*9.96*	*9.72*	*5.31*	*16.41*	*15.41*	*0.36*
	Simon	BF_{10}	**0.08**	**0.34**	**0.07**	**0.14**	**0.87**	**0.11**	**0.49**	**0.14**	**0.09**
		BF_{01}	*12.86*	*2.93*	*14.03*	*7.39*	*1.15*	*8.72*	*2.03*	*7.19*	*11.65*
	Anti	BF_{10}	**0.76**	**0.09**	**0.15**	**0.06**	**0.10**	**0.06**	**0.08**	**0.06**	**0.19**
	-saccade	BF_{01}	*1.31*	*10.96*	*6.89*	*16.34*	*9.61*	*16.62*	*12.77*	*16.46*	*5.19*

Note. For the sake of clarity, BF_{10} are presented in bold, whereas BF_{01} are presented in italics. BF = Bayes Factor; num. = numerical;

upd. = updating; c.s. = complex span.

Appendix G

Bayes Factors for the Pearson Correlation Coefficients

Table G1

Bayes Factors in favor of the Alternative Hypothesis (BF_{10}) and in favor of the Null hypothesis (BF_{01}) for the Pearson Correlation Coefficients.

	BF	Color Stroop	Number Stroop	Arrow flanker	Letter flanker	Simon	Antisaccade	Stop-signal
Number Stroop	BF_{10}	**0.17**	-					
	BF_{01}	*5.90*						
Arrow flanker	BF_{10}	**0.07**	**0.06**	-				
	BF_{01}	*15.20*	*16.89*					
Letter flanker	BF_{10}	**0.10**	**0.07**	**0.12**	-			
	BF_{01}	*10.22*	*14.61*	*8.08*				
Simon	BF_{10}	**0.19**	**0.16**	**0.09**	**0.06**	-		
	BF_{01}	*5.31*	*6.22*	*11.28*	*16.95*			
Antisaccade	BF_{10}	**0.07**	**0.06**	**0.07**	**0.06**	**0.53**	-	
	BF_{01}	*13.79*	*16.89*	*14.72*	*16.68*	*1.87*		
Stop-signal	BF_{10}	**0.07**	**0.07**	**0.16**	**0.12**	**0.40**	**0.08**	-
	BF_{01}	*14.80*	*13.98*	*6.39*	*8.30*	*2.53*	*12.98*	
Letter sets	BF_{10}	**0.08**	**0.06**	**0.13**	**4.05**	**0.07**	**0.06**	**0.07**
	BF_{01}	*11.91*	*15.93*	*7.94*	*0.25*	*14.61*	*16.85*	*14.81*
Locations	BF_{10}	**0.10**	**0.13**	**0.12**	**0.10**	**0.08**	**0.07**	**0.56**
	BF_{01}	*10.18*	*7.89*	*8.52*	*9.74*	*12.81*	*15.22*	*1.78*
RAPM	BF_{10}	**0.06**	**0.07**	**0.09**	**0.87**	**0.48**	**0.08**	**0.06**
	BF_{01}	*16.55*	*14.14*	*11.63*	*1.15*	*2.10*	*12.17*	*15.60*
Relationships	BF_{10}	**0.10**	**0.06**	**0.36**	**0.29**	**0.06**	**0.07**	**0.06**
	BF_{01}	*9.88*	*16.86*	*2.76*	*3.42*	*15.85*	*13.46*	*16.88*

	BF	Color Stroop	Number Stroop	Arrow flanker	Letter flanker	Simon	Antisaccade	Stop-signal
Syllogisms	BF_{10}	**0.06**	**0.06**	**0.06**	**0.07**	**0.06**	**0.06**	**0.11**
	BF_{01}	*16.60*	*16.36*	*16.51*	*14.01*	*16.69*	*16.50*	*8.81*
Numerical updating	BF_{10}	**0.06**	**0.06**	**0.09**	**0.06**	**0.07**	**0.07**	**0.07**
	BF_{01}	*16.67*	*15.97*	*11.05*	*15.73*	*15.33*	*13.80*	*13.70*
Spatial updating	BF_{10}	**0.06**	**0.06**	**0.07**	**0.45**	**2.59**	**0.07**	**0.07**
	BF_{01}	*16.63*	*16.90*	*14.72*	*2.22*	*0.39*	*14.83*	*14.68*
Numerical complex span	BF_{10}	**0.18**	**0.16**	**0.11**	**0.11**	**0.06**	**0.09**	**0.08**
	BF_{01}	*5.49*	*6.22*	*8.73*	*9.21*	*16.24*	*11.68*	*13.10*
Spatial complex span	BF_{10}	**0.17**	**0.09**	**0.07**	**0.10**	**0.08**	**0.06**	**0.38**
	BF_{01}	*6.01*	*11.66*	*15.14*	*9.74*	*12.82*	*16.95*	*2.64*
Time estimation	BF_{10}	**0.06**	**0.08**	**1.57**	**0.06**	**0.37**	**0.18**	**0.06**
with timer	BF_{01}	*16.39*	*13.22*	*0.64*	*16.92*	*2.73*	*5.43*	*16.86*
Time estimation	BF_{10}	**0.15**	**0.17**	**0.06**	**0.06**	**0.06**	**0.08**	**0.24**
without timer	BF_{01}	*6.60*	*6.05*	*16.85*	*16.52*	*16.80*	*12.24*	*4.25*

(Table F1 continues)

		Letter sets	Locations	RAPM	Relation ships	Syllogisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.c.	Timer
Locations	BF_{10}	**4.65e+04**	-								
	BF_{01}	*2.15e-05*									
RAPM	BF_{10}	**1.78e+07**	**1.45e+03**	-							
	BF_{01}	*5.62e-08*	*6.90e-04*								
Relationships	BF_{10}	**4.65e+12**	**2.66e+05**	**1.72e+03**	-						
	BF_{01}	*2.15e-13*	*3.76e-06*	*5.81e-04*							
Syllogisms	BF_{10}	**0.62**	**2.10e+03**	**0.15**	**431.55**	-					
	BF_{01}	*1.60*	*4.75e-04*	*6.47*	*2.32e-03*						

		Letter sets	Locations	RAPM	Relation ships	Syllogisms	Num. upd.	Spatial upd.	Num. c.s.	Spatial c.c.	Timer
Num.	BF_{10}	**1.51e+07**	**551.58**	**2.17e+03**	**2.89e+03**	**1.59**	-				
upd.	BF_{01}	*6.62e-08*	*1.81e-03*	*4.61e-04*	*3.46e-04*	*0.63*					
Spatial	BF_{10}	**8.74e+06**	**1.50e+05**	**101.77**	**1.45e+09**	**236.32**	**7.61e+07**	-			
upd.	BF_{01}	*1.14e-07*	*6.68e-06*	*0.01*	*6.89e-10*	*4.23e-03*	*1.31e-08*				
Num.	BF_{10}	**21.11**	**0.22**	**8.81**	**2.02**	**0.06**	**8.80e+05**	**131.58**	-		
c.s.	BF_{01}	*0.05*	*4.59*	*0.11*	*0.50*	*16.76*	*1.14e-06*	*0.01*			
Spatial	BF_{10}	**2.25e+04**	**305.45**	**158.26**	**1.27e+03**	**0.13**	**3.50e+09**	**7.75e+08**	**2.74e+09**	-	
c.s.	BF_{01}	*4.45e-05*	*3.27e-03*	*0.01*	*7.86e-04*	*7.55*	*2.86e-10*	*1.29e-09*	*3.65e-10*		
Time estimation	BF_{10}	**0.07**	**0.09**	**0.36**	**0.07**	**0.11**	**0.63**	**0.11**	**0.06**	**0.38**	-
with timer	BF_{01}	*14.18*	*10.99*	*2.75*	*14.06*	*8.72*	*1.59*	*9.39*	*15.50*	*2.64*	
Time estimation	BF_{10}	**0.90**	**84.25**	**1.19**	**3.07**	**0.22**	**268.04**	**45.20**	**2.24**	**9.54e+06**	**0.07**
without timer	BF_{01}	*1.11*	*0.01*	*0.84*	*0.33*	*4.49*	*3.73e-03*	*0.02*	*0.45*	*1.05e-07*	*14.13*

Note. For the sake of clarity, BF_{10} are presented in bold, whereas BF_{01} are presented in italics. BF = Bayes Factor; num. = numerical; upd. = updating; c.s. = complex span.