# *A true reflection of executive functioning or a representation of task-specific variance? Re-evaluating the unity/diversity framework*

Acta Psychologica 236 (2023) 103934

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

# A true reflection of executive functioning or a representation of task-specific variance? Re-evaluating the unity/diversity framework

Stjepan Sambol [*], Emra Suleyman, Jessica Scarfo, Michelle Ball

*Institute for Health and Sport (IHES), Victoria University, Melbourne, Victoria, Australia*

ARTICLE INFO

*Keywords:*
Executive function
Factor analysis
Unity/diversity framework
Updating
Shifting
Inhibition

ABSTRACT

The unity/diversity framework, originally published by Miyake et al. (2000) has become the most cited model of executive functioning. Consequently, when researchers operationalise executive function (EF) they often exclusively assess the three "core" EFs: updating, shifting, and inhibition. However, rather than core EFs representing domain general cognitive abilities, these three EFs may instead represent specific procedural skills from the overlapping methodologies of the tasks selected. In this study, we conducted a confirmatory factor analysis (CFA) which showed both the traditional three-factor and nested-factor model from the unity/diversity framework failed to reach satisfactory levels of fit. Subsequently, an exploratory factor analysis supported a three-factor model reflecting: an expanded working memory factor, a combined shifting/inhibition factor representing cognitive flexibility, and a factor comprised solely of the Stroop task. These results demonstrate that working memory remains the most robustly operationalised EF construct, whereas shifting and inhibition may represent task-specific mechanisms of a broader domain-general cognitive flexibility factor. Ultimately, there is little evidence to suggest that updating, shifting, and inhibition encapsulates all core EFs. Further research is needed to develop an ecologically valid model of executive functioning that captures the cognitive abilities associated with real world goal-directed behaviour.

## 1. Introduction

Historically labelled "frontal lobe functions", there has been extensive research investigating executive function (EF), with escalating interest since the start of the twenty-first century (Baggetta & Alexander, 2016; Karr et al., 2018). Over the years, many definitions and models regarding the nature and organisation of EF have been proposed. For example, Goldstein and Naglieri (2014) identified 32 separate definitions of EF within the literature. Nonetheless, there is a considerable degree of overlap across these models and definitions, and leading researchers have commonly – although not unanimously – categorised EF as an umbrella term that encompasses a variety of distinct but related higher order cognitive control processes. These control processes, labelled executive functions (EFs), presumably coordinate lower-level cognitive processes (i.e., attention, language, memory) to facilitate intentional self-regulation and goal-directed or problem-solving behaviours (i.e., executive functioning; Diamond, 2013; Friedman & Miyake, 2017; Lezak et al., 2004). Accordingly, EF can be thought of as analogous to the conductor of an orchestra, coordinating multiple different components to create a coherent and desirable symphony (Jones et al., 2012).

To date the most cited model of EF is the unity/diversity framework (Miyake et al., 2000; 17,790 citations according to Google Scholar as of February 2023). Following an extensive literature review Miyake et al. (2000) proposed that self-regulatory and goal directed behaviour was influenced by three core EFs: shifting attention in response to environmental change, monitoring and updating the content of working memory, and inhibition of impulses and inappropriate responses (i.e., shifting, updating, inhibition). A test battery was designed to assess each of the three EFs, which comprised nine computerised tasks; three for each EF (Miyake et al., 2000). Miyake et al. (2000) conducted a confirmatory factor analysis (CFA), which was stipulated to extrapolate a "purer" measure of each executive construct as the shared variance would isolate the desired EF and alleviate the issue of task impurity. Often performance on an EF measure is influenced by multiple higher-order and lower-order cognitive processes acting simultaneously. Therefore, it is difficult to obtain a score from a task that isolates an EF (i.e., task impurity). The CFA supported a three-factor model, showing

* Corresponding author.
*E-mail address:* Stefan.sambol@vu.edu.au (S. Sambol).

https://doi.org/10.1016/j.actpsy.2023.103934
Received 28 February 2023; Received in revised form 21 April 2023; Accepted 4 May 2023
Available online 6 May 2023

shifting, updating, and inhibition to be separable EFs (Miyake et al., 2000). However, the three EFs were not clearly distinguishable constructs as they were moderately correlated with one another, demonstrating the "unity" and "diversity" between these EFs (Miyake et al., 2000). Shifting, updating, and inhibition were also found to be significant predictors of performance on complex EF tasks such as the Tower of London and Wisconsin Card Sorting Task; reaffirming the notion that they are core EFs (Miyake et al., 2000). Consequently, the unity/diversity framework re-conceptualised EF as comprising a dissociable but interrelated set of cognitive processes and showed that CFA is an effective technique to determine the organisation of EF (Karr et al., 2018).

The original three-factor model proposed by Miyake has been modified to a "nested-factor" model (Friedman & Miyake, 2017; Miyake & Friedman, 2012). Rather than including three separate EFs as factors, the nested-factor model introduces a "common" EF factor, which encapsulates the shared variance between updating, shifting, and inhibition (see Fig. 1). The common EF factor potentially reflects a domain-general executive control construct, which emerges in early childhood (Best & Miller, 2010) and is the foundational to successful performance on all EF measures and goal-directed behaviour (i.e., executive functioning) more broadly (von Bastian et al., 2020). This construct represents an individual's ability to bias attention towards a particular goal or task, the capacity to create mental representations of that goal, and to coordinate lower-order cognitive abilities to support in achieving that goal (Friedman & Miyake, 2017; Miyake & Friedman, 2012). Once the shared variance between measures of updating, shifting, and inhibition were accounted for in the nested-factor model, there was no unique variance remaining for a distinct inhibition factor with the unaccounted variance forming both updating-specific and shifting-specific factors (Friedman & Miyake, 2017).

There has been extensive research investigating the unity/diversity framework across all age groups. During the earliest stage of life (i.e., infancy; 0–2 years), EF appears to be best represented as a unitary construct (Wiebe & Karbach, 2018). Undoubtedly, this is due to the considerable difficulties in measuring EF during infancy. This age period coincides with extensive neurological development and limitations in motor, language, and attention abilities (Hughes & Graham, 2002; Isquith et al., 2005). As children age, executive functioning continues to improve with children beginning to show signs of updating working memory, shifting attention, and basic response inhibition as early as age four (Best & Miller, 2010). Subsequently, research into models of EF during childhood (4–12 years) primarily supported a unitary factor model of EF, or a two-factor model represented by a distinct updating factor, and combined shifting and inhibition factor (Brocki & Tillman, 2014; Brydges et al., 2012; Lambek & Shevlin, 2011; Lee et al., 2012; Lehto et al., 2003; Monette et al., 2015; Usai et al., 2014). The lack of a distinct inhibition and shifting factor has been attributed to children possessing domain-general cognitive processes rather than executive processes that are attuned to specific task requirements (Wiebe & Karbach, 2018).

By adolescence (12–18 years), and adulthood research has shown comparable support for both the three-factor (Fournier-vicente et al., 2008; Klauer et al., 2010; Lee et al., 2013; Xu et al., 2013) and nested-factor models (Fleming et al., 2016; Ito et al., 2015; Miyake & Friedman, 2012). However, this is not uniformly the case, with other adult studies supporting a two-factor model with a distinct updating factor and combined shifting and inhibition factor (Adrover-Roig et al., 2012; Bettcher et al., 2016; Hedden & Yoon, 2006; Hull et al., 2008). A systematic review by Karr et al. (2018) which examined the unity/diversity framework only found tentative support for the model. Upon conducting a bootstrap re-analysis which simulated 5000 samples, no model was unequivocally supported. Across all age groups, the three- and nested-factor models showed low rates of model acceptance (i.e., <50 % of the time models met conventional fit thresholds; Karr et al., 2018). Overall, although a portion of the early literature has shown some degree of support for the unity/diversity framework, emerging research has challenged Miyake et al.'s (2000) conceptualisation of EF.

Despite being the most cited model of executive functioning, the unity/diversity framework has several important limitations, which are often repeated in replications throughout the research literature. Firstly, across numerous studies the inhibition tasks used have demonstrated weak factor loadings and were unable to form a unitary latent variable (Brydges et al., 2012; Hedden & Yoon, 2006; Hull et al., 2008; Rey-Mermet et al., 2019). Miyake and Friedman (2012) theorised that the ability to inhibit irrelevant information or prepotent responses may be entirely captured by a unitary cognitive control network (i.e., common EF), rather than an inhibition-specific process. However, the transition from the original three-factor model to the nested-factor model fails to alleviate this issue as studies are unable to achieve satisfactory levels of fit with the nested factor model (Karr et al., 2018). Problematically, the inhibition measures used to construct the unity/diversity framework (e.g., antisaccade task, flanker task, Stroop, go/no-go, and the stop-signal task) have demonstrated weak-to-no relationships between each other

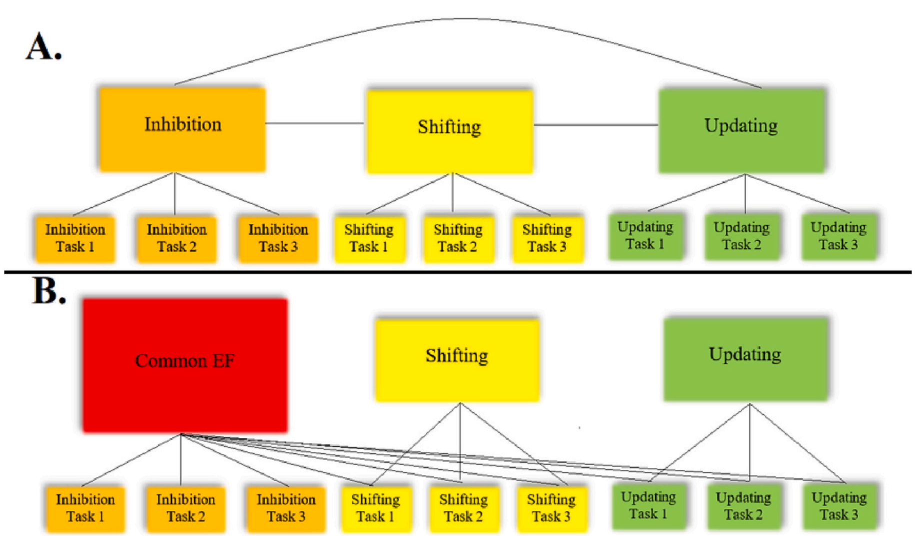


**Fig. 1.** Three-factor and Nested-factor model of the Unity/Diversity Framework.
*Notes.* Model A represents the three-factor model with distinct yet related inhibition, shifting, and updating factors. Model B represents the nested factor model with a common EF, shifting specific, and updating specific factor.

(Guye & von Bastian, 2017; Hedge et al., 2018;Rey-Mermet et al., 2019 ; Rouder & Haaf, 2019 ; von Bastian et al., 2020).

A recent wave of investigation into the psychometric properties of the aforementioned inhibition measures (occasionally labelled as "executive" or "attentional control" measures) has resulted in two competing explanations regarding the weak correlations between tasks. Some researchers have stipulated that there is no unitary inhibition factor, but rather multiple domain-specific inhibition skills divided by task mechanics or the type of distractors implemented (Rey-Mermet et al., 2019; Rouder & Haaf, 2019; von Bastian et al., 2020). Conversely, other researchers have posited that the tasks are theoretically assessing the same construct (Draheim et al., 2021; Hedge et al., 2018), although this construct may operate as a higher-order factor that can be divided in lower-order factors based on task mechanics or the type of distractors (von Bastian et al., 2020). Von Bastian et al. (2020) attributed the lack of convergent validity between inhibition measures to poor reliability properties.

Hedge et al. (2018) examined the intraclass correlations of the flanker task, Stroop, go/no-go, and the stop-signal task, and most measures fell below acceptable reliability criteria. The poor reliability characteristics of these tasks were credited to the use of reaction time as the dependent variable without accounting for accuracy (von Bastian et al., 2020). The scoring method most commonly used for these inhibition measures is to compare reaction time on experimental conditions (typically containing added distractions) with control conditions. However, this scoring method fails to account for participants' speed-accuracy trade-off during experimental conditions, where some participants may favour accuracy over speed or vice-versa. Additionally, the reaction time difference assumes the influence of processing speed is consistent across the different trials, yet there is no evidence for this assumption (Draheim et al., 2021; Rey-Mermet et al., 2019). The unintended influence of speed-accuracy trade-off and processing speed increases the unsystematic variance of inhibition scores and ultimately diminishes the reliability of these tasks. Subsequently, the poor reliability properties are problematic when assessing correlations between constructs as the results might be attributable to error (Draheim et al., 2021; Hedge et al., 2018; Rey-Mermet et al., 2019). Draheim et al. (2021) showed that relationships between the Stroop and flanker tasks improved when accuracy-based dependent variables were used instead of reaction time. Furthermore, the tasks showed stronger reliability properties and also formed a coherent inhibition latent factor, adding support for a unitary inhibition factor.

Collectively, these recent studies call into question what precise cognitive ability the inhibition factor from the unity/diversity framework represents, given the inconsistent relationships between inhibition tasks and their poor reliability properties. The inhibition factor should be reconstructed with tasks that avoid these limitations. For example, the Victoria Stroop test, contingency naming test (CNT), and stop-signal task (SST) are measures of inhibition that factor both speed and accuracy into the inhibition scores provided, and demonstrate sound reliability properties (see Anderson et al., 2000; Strauss et al., 2006; von Bastian et al., 2020). These alternative tasks can be used to determine whether a unitary inhibition latent factor can be constructed or if inhibition is best represented through multiple domain-specific factors.

The second limitation with the unity/diversity framework is that previous studies which have supported the shifting factor utilised tasks with procedurally identical methodology. These studies administered three task-switching paradigms (e.g., The colour-shape, letter-numbers, picture-symbol, and category-switch) to construct the shifting factor (Fleming et al., 2016; Fournier-vicente et al., 2008; Friedman et al., 2011, 2016; Ito et al., 2015; Klauer et al., 2010; Lee et al., 2013; Xu et al., 2013). Although conceptually these tasks are different, they are procedurally identical with only superficial changes to the stimuli presented (e.g., presenting a red/green triangles/circles compared to presenting a vowel/consonant with an even/odd number, see Fig. 2).

When conducting a factor analysis with multiple task-switching paradigms, the resulting latent variable is likely to reflect more than just shifting abilities due to the overlapping methodology. For example, while task switching paradigms require participants to adapt to new rule changes, they also involve filtering out distracting information, fine motor-skills (i.e., the procedural skill of alternating between two fixed keyboard responses), and processing speed (Yu et al., 2017). The constructed shifting factor from the unity/diversity framework could just as easily represent these abilities, rather than isolating shifting. To truly construct a shifting factor it is necessary to select tasks which require adjusting attention to new rule changes, while avoiding overlap in assessing undesired abilities. For example, alongside a task-switching paradigm researchers could also administer card sorting tasks which assess participants' ability to adapt to rule changes without being influenced by processing speed, as these tasks are not time-based (Perry et al., 2001). Additionally, verbal fluency is not often associated with shifting, but has been shown to significantly relate to task switching paradigms and measures participants' ability to generate new information (Paula et al., 2015), without concern that performance will be constrained by their ability to inhibit distractions caused by learned behaviour (e.g., responding to the colour rather than shape of a stimulus).

Overall, the unity/diversity framework has been widely replicated in the research literature, becoming the most prominent model of EF, however these studies often repeated the same limitations outlined

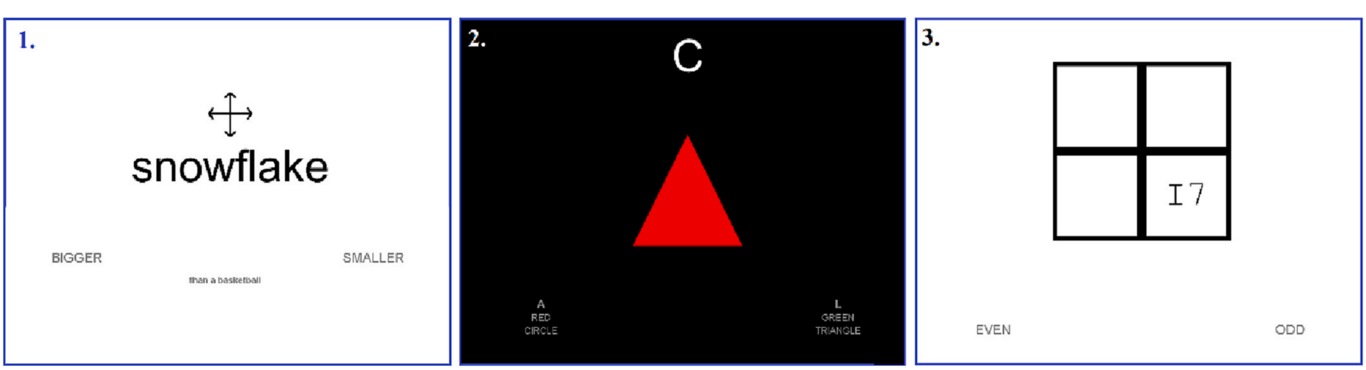


**Fig. 2.** Three task-switching measures: category-switch, colour-shape, and letter-numbers.

*Notes.* Photo 1 is a screen capture of the category-switch task, participants alternate between identifying if the object named in the centre of the screen is living/non-living, or smaller/larger than a basketball. Photo 2 presents the colour-shape task where participants identify the shape or colour of the object presents based on the letter cue at the top of the screen. Photo 3 depicts the letters-numbers task, when the letter/number compound is presented in the lower quadrant participants identify the number (i.e., odd/even) and when the letter/number compound are presented in the upper quadrant participants identify the letter (i.e., value or consonant).

above: selecting unreliable inhibition measures and multiple task-switching measures with overlapping methodologies, thus diminishing the validity of the unity/diversity framework. It is important that EF is operationalised and measured accurately as executive functioning has been associated with numerous positive life outcomes (e.g., academic ability, physical health, personal finances, social interactions; Alloway & Alloway, 2010; Moffitt et al., 2011). Furthermore, there is clinical utility in ensuring popularised models of EF are sufficiently validated as executive dysfunction has been associated with most psychological disorders (Snyder et al., 2015). Therefore, models of EF that can accurately identify the core EFs can assist clinicians with selecting which tasks to use when screening for deficits in executive functioning. This study aimed to explore the relationship between nine EF measures and their associated latent factors, firstly, by conducting confirmatory factor analyses to re-evaluate the three-factor and nested-factor model of the unity/diversity framework, and secondly, by carrying out an exploratory factor analysis (EFA).

The EF test library is immense, Baggetta and Alexander (2016) reviewed 106 studies and identified 109 distinct tasks and 11 different test batteries proclaiming to measure 39 uniquely labelled EFs. Studies which have supported the unity/diversity framework have intentionally selected shifting tasks where there is little change in the procedural demands for participants. Furthermore, inhibition measures used have shown problematic reliability properties (Rey-Mermet et al., 2019; von Bastian et al., 2020). The tasks selected in the current study are purported to measure factors from the unity/diversity framework, but with no overlapping procedures and demonstrated sound psychometric properties. By selecting measures which theoretically assess the same cognitive construct through varied procedures, there should be an increased likelihood that latent factors represent a specific EF, rather than a task-specific procedural skill.

## 2. Method

### 2.1. Participants

A sample of 156 participants aged between 18 and 58 years, were recruited through convenience and snowball sampling methods via electronic advertisements posted on social media (e.g., Facebook) and on university notice boards. Inclusion criteria for the current study required participants aged between 18 and 60 years, normal or corrected to normal vision, and no currently diagnosed psychological disorder (e.g., mood disorders, anxiety disorders, etc.) or pre-existing neurological disorder (e.g., Alzheimer's and Dementia, ADHD, etc.). Demographic information is presented in Table 1. For taking part in the study each participant was compensated with a $20 gift card and an optional brief personality report (that was produced as part of a larger study).

**Table 1**
Participant demographic information ($N = 156$).

| Variable | *M*(*SD*) | *n*(%) |
|---|---|---|
| Gender | | |
| Male | | 66 (42.3 %) |
| Female | | 90 (57.7 %) |
| Age | 25.92 (7.47) | |
| Education | | |
| High School | | 54 (34.6 %) |
| Certificate 3 or 4 | | 17 (10.9 %) |
| Diploma or Advanced Diploma | | 13 (8.3 %) |
| Bachelor or Honours Degree | | 69 (44.2 %) |
| Master's Degree | | 1 (0.6 %) |
| Doctorate | | 1 (0.6 %) |

### 2.2. Materials

A demographics questionnaire was administered via the online platform Inquisit, collecting information on participants' eligibility to participate, age, gender, education, current occupation, and occupation classification. Additionally, an extensive EF test battery was administered including three measures for each EF construct: updating, shifting, and inhibition. A comparison between the tasks used in the current study and the tasks used by Miyake et al. (2000) and Friedman et al. (2008) is presented in Table 2. Table 3 presents a summary of the psychometric properties for each EF tasks used in the current study.

#### 2.2.1. Updating measures

In keeping with Miyake et al.'s (2000) definition, updating refers to monitoring and manipulating the content of working memory. The digit span task, n-back, and keep track were selected as measures of updating.

*2.2.1.1. Digit span task.* The experimenter read a list of digits to participants, at a rate of one digit per second (Wahlstrom et al., 2016). For digit span forward (DF), the participants recited the digits in the order presented, whereas for digit span backwards (DB) digits were recited in reverse order. Both DF and DB began at a digit length of two with increasingly longer sequences after every two trials to a maximum of nine and eight digits, respectively (Wechsler, 1997). Participants received one point for each correctly recited DB trial, if a mistake was made on two consecutive trials of the same digit length the test was discontinued.

*2.2.1.2. N-back task.* Participants were presented with a sequence of $20 + n$ yellow letters (c, g, h, k, p, q, t, w) against a black background, where $n$ represents the current level of the task (i.e., 2-, 3-, or 4-back; Jaeggi et al., 2010a, 2010b). Each letter was presented for 500 ms, followed by a 2500 ms interstimulus interval. Participants were tasked with identifying whether the current letter matches the letter presented two- (2-back), three- (3-back), or four- stimulus prior (4-back). Overall, this n-back task consisted of nine trials, three for each n-back level. The difference between the total number of correct hits and commission errors (i.e., incorrect identification of target), divided by the number of blocks was used as a measure of updating (Jaeggi et al., 2010a, 2010b).

*2.2.1.3. Keep track.* The keep track task comprised of 12 trials (Miyake et al., 2000), and each presented participants with 2–4 key categories, followed by a list of 15 words presented serially at the centre of a computer screen for 1500 ms each, in a randomised order. At the end of the 15-word sequence participants were asked to recall the last word presented for each of the key categories presented at the start of each trial. Initial trials began with only two key categories and increased in difficulty every four trials by adding an additional key category up to a

**Table 2**
Comparison of EF measures used in the current study and past research.

| Current study | Miyake et al., 2000 | Friedman et al., 2008 |
|---|---|---|
| Updating | | |
| Digit Span | Letter Memory | Letter Memory |
| N-back | Tone Monitoring | Spatial n-back |
| Keep Track | Keep Track | Keep Track |
| Shifting | | |
| Colour-shape | Colour-shape | Plus-minus |
| MCST | Number-letter | Number-letter |
| Verbal Fluency | Category-switch | Local-global |
| Inhibition | | |
| Stroop | Stroop | Stroop |
| Stop-signal – integration method | Stop-signal – mean method | Stop-signal – mean method |
| CNT | Antisaccade | Antisaccade |

*Notes.* CNT = Contingency naming test; MCST = Modified card sorting task.

**Table 3**
Psychometric properties of EF measures used in the current study.

| EF Measure | Validity | Reliability |
|---|---|---|
| Digit Span | Moderate correlations with other measures of working memory: Brown-Peterson task ($r$ = 0.41; Geurten et al., 2016) and Digit Symbol test ($r$ = 0.32; Woods et al., 2011). | Strong test-retest reliability ($r$ = 0.64–0.84; Woods et al., 2011). |
| N-Back | Correlated with change detection task used to assess working memory capacity ($r$ = 0.40; Frost et al., 2021). Spatial n-back showed factor loadings >0.50 with other updating measures (Friedman et al., 2016). | Strong test-retest reliability ($r$ = 0.79; Soveri et al., 2018). |
| Keep Track | Factor loadings >0.50 with other measures of updating (Friedman et al., 2016). | Sound internal consistency ($a$ = 0.65; Friedman et al., 2016) and odd-even reliability ($r$ = 0.60; Ito et al., 2015). |
| CST | Factor loadings >0.50 with other measures of shifting (Friedman et al., 2016).[a] | Sound split-half reliability ($r$ = 0.85; Friedman et al., 2016) and odd-even reliability ($r$ = 0.88; Ito et al., 2015). |
| MCST | Correlations with other shifting measures ranging between $r$ = 0.13–0.36 (Lineweaver et al., 1999). WCST showed factor loading of 0.50 with other shifting measures (Hedden & Yoon, 2006). | Sound test-retest reliability ($r$ = 0.64; Kopp et al., 2019b) and split half reliability $r$ > 0.90 (Kopp et al., 2019a). |
| Verbal Fluency | Factor loading of 0.50 with the trail making test (Lehto et al., 2003), and significantly correlated with the Wisconsin card sorting task ($r$ = 0.22; Brydges et al., 2012). | Sound test-retest reliability of ($r$ = 0.74) and internal consistency (α =0.83; Tombaugh et al., 1999). |
| EF Measure | Validity | Reliability |
| Stroop | Weak relationship observed with the *ReacStick* test of behaviour inhibition ($r$ = 0.20; van Schooten et al., 2019).[b] | High test-retest reliability coefficients for all three parts of the test ($r$ = 0.90, $r$ = 0.83, $r$ = 0.91; Strauss et al., 2006). |
| SST | Factor loading of 0.50 with other measures of inhibition (Gunten et al., 2020). | Sound split-half reliability ($r$ = 0.97; Gunten et al., 2020) |
| CNT | Factor loadings >0.40 with other measures of inhibition (Riddle & Suhr, 2012; Testa et al., 2012). | Moderate to strong test-retest correlations for completion time on all trials ($r$ = 0.82, $r$ = 0.62, $r$ = 0.46, and $r$ = 0.59; Riddle & Suhr, 2012).[c] |

Notes. CNT = Contingency naming test; CST = Colour-shape task; EF = Executive function; MCST = Modified card sorting task; SST = Stop-signal time; Wisconsin card sorting task.
[a] Other measures of shifting only include other task-switching paradigms. [b] To-date no study has investigated the relationship between Victoria Stroop test and other commonly administered inhibition measures. [c] Test-retest conducted over a two week period with only 26 participants.

maximum of four key categories. The total number of correctly recalled words provided a measure of updating.

#### 2.2.2. *Shifting measures*

Shifting is the ability to switch between tasks or mental sets (Miyake et al., 2000) and was measured using the colour-shape task, modified card sorting task, and verbal fluency.

*2.2.2.1. Colour-shape task (CST).* The CST consisted of a practice block with 24 trials and two experimental blocks with 56 trials each (Friedman et al., 2016, 2018). For each trial, a cue is presented (the letter 'C' or 'S') followed by a shape (square or triangle) in one of two colours (green or red). Participants were tasked with categorising the stimuli by colour or shape, depending on the letter cues, by pressing 'A' or 'L' on the keyboard as quickly and accurately as possible. The difference between the average reaction time (RT) on switch trials and repeat trials (i.e., the switch cost) was used as a measure of participants' shifting ability.

*2.2.2.2. Modified card sorting task (MCST).* Participants were presented with four reference cards and a single card from two decks of 24 cards. Cards from the decks varied in terms of colour (red, green, blue, or yellow), shape (circle, square, star, or cross), and/or number (1, 2, 3, or 4). Participants were tasked with sorting each of the 48 cards with one of the four reference cards. The sorting rule cycled through a pre-determined sequence (colour, shape, number) until all 48 cards have been played, with the sorting rule changing after six consecutive correct answers, completing the current category. The total number of perseverative errors (i.e., failure to change, and continued use of a sorting rule following feedback that it was incorrect) provided a measure of the participants' shifting ability. Notably the current study followed the Channon (1996) scoring method for perseverative errors, however also scored alternating between two incorrect sorting rules as perseverative errors (i.e., colour, shape, colour, shape).

*2.2.2.3. Verbal fluency.* Participants were asked to produce within one minute as many different words as they could for a given letter: F', 'A' and 'S' (Borkowski et al., 1967; Strauss et al., 2006). Participants were instructed to avoiding repeating words or saying proper-nouns. All participants' spoken words were written down and the total number of unique words produced after the three one-minute periods was used as a measure of shifting (Paula et al., 2015). Verbal fluency is not often used as a measure of shifting, however, recent research showed performance on the F, A, S verbal fluency test significantly correlated with the shifting attention test ($r$ = 0.23; Aita et al., 2019).

#### 2.2.3. *Inhibition measures*

Inhibition refers to the inhibiting automatic responses and resisting distraction from irrelevant stimuli (Miyake et al., 2000). Inhibition was measured with three measures: Stroop test, stop-signal task, and contingency naming test.

*2.2.3.1. Stroop test.* This study used a computerised version of the Victoria Stroop Test, which included three cards (one for each trial) each with twenty-four stimuli arranged in a 6 × 4 matrix (Strauss et al., 2006). The stimuli were presented in one of four colours: blue, green, red or yellow. In Trial 1, coloured rectangles were presented, and in Trial 2, the rectangles were replaced with simple words (e.g., hard, when), and in Trial 3, colour-words were presented (e.g., blue, green, red or yellow) in an incongruent colour (e.g., "blue" written in red ink). In each trial, starting with the top row and working from left to right, participants were asked to name the colour of the stimuli as quickly as possible. Trial 3 assesses participants' ability to supress a habitual response (i.e., reading the colour word) and produce a less familiar one (i.e., naming the incongruent colour) while simultaneously inhibiting potential distractions from adjacent items. Therefore, a measure of inhibition control was ascertained by calculating the ratio between the time to complete Trial 3 and time to complete Trial 2.

*2.2.3.2. Stop-signal task (SST).* The SST consisted of one practice block of 20 trials (5 stop-signal trials; 15 no-signal trials = 1:3 ratio), and three blocks of 68 trials (17 stop-signal trials; 51 no-signal trials = 1:3 ratio). At the beginning of each trial a fixation point is presented at the centre of the screen. During no-signal trials, the fixation point is followed by an arrowhead (left or right) and participants were tasked with pressing "D" for left arrowheads and "K" for right arrowheads on the keyboard as quickly as possible. Stop-signal trials followed a similar procedure except participants were required to cease any action if a signal beep was played after the presentation of the arrow. The delay between the presentation of the arrow and signal beep on stop-signal trials started at 250 ms and would increase or decrease by 50 ms, to a maximum delay of 1150 ms or minimum delay of 50 ms, for each unsuccessful and successful stop-signal trial, respectively. An estimated stop-signal reaction time (SSRT) was calculated using the new integration method shown to produce the most reliable and least biased result (Verbruggen et al., 2019).

*2.2.3.3. Contingency naming test (CNT).* This study's computerised CNT consisted of four trials where participants were presented with a card on screen containing 27 items arranged in a 3 × 9 matrix. Each item comprises of an outer shape (circle, square, or triangle) and inner shape (circle, square, or triangle), coloured in either: blue, green, or pink. In Trial 1, participants were tasked with naming the colour of each item, and in Trial 2, participants were asked to name the outer shape. In Trial 3, participants were tasked with naming the colour of the item when the inner and outer shapes are congruent, otherwise they named the outer shape. Lastly, Trial 4 maintains the same naming rule as Trial 3, although when an arrow was presented above an item the rule was reversed (Anderson et al., 2000). A practice trial using a nine-item card was completed before each of the four experimental trials. A measure of inhibition control was ascertained by summing the efficiency score for Trial 3 and 4. The efficiency score takes into account completion time and number of errors (calculated using the formula $[(1/\text{time}) / \sqrt{(\text{errors}+1)}] \times 100$; see Anderson et al., 2000). Trials 1 and 2 assess non-executive cognitive abilities such as general processing speed, whereas Trials 3 and 4 involve more complex naming rules requiring individuals to supress previously learned rules and produce a less familiar response. Therefore, a greater efficiency score on Trial 3 and 4 indicates superior inhibition control (Riddle & Suhr, 2012).

### 2.3. Procedure

The current study received ethics approval from the relevant Human Research Ethics Committee. An online meeting over Zoom® was arranged via email with each participant. Upon meeting online, participants verbally confirmed informed consent and then completed tests of updating (digit span, n-back, keep track), shifting (CST, MCST, verbal fluency) and inhibition (Stroop, SST, CNT). All computerised tests (n-back, keep track, CST, MCST, Stroop, SST, and CNT) were accessed through Inquisit® (a web-based program which hosts cognitive tests) links provided during the meeting and were required to share screen. The EF testing session took approximately 90 min to complete with two five-minute breaks imbedded into the testing session. Lastly, the EF tests were counterbalanced, with participants completing the tests in one of three predetermined orders.

### 2.4. Statistical design

Test performance data was imputed, cleaned, and analysed using IBM® SPSS® Statistics Version 27. Preliminary analyses involved obtaining descriptive statistics and Pearson's correlations for each variable assessed.

The primary analyses involved replicating the three-factor and nested-factor models of the unity/Diversity framework using CFA with Amos® Version 25. During data screening, values beyond three standard deviations from the mean were replaced with values three standard deviations from the mean for each variable to minimise the effect of outliers. This was the same procedure for handling outliers that was used by Friedman et al. (2008). Secondly, two participants were removed for failing to complete all cognitive tests. A further eight participants were removed for average reaction times slower than 2000 ms or faster than 200 ms on the CST as this jeopardises the switch cost calculation (Sicard et al., 2022). Lastly, another 12 participants were removed for either violating the deviation assumption (i.e., accuracy on stop-signal trials deviated beyond 25 %–75 %) or the race model assumption (i.e., stop-signal trials reaction time is greater than reaction time on non-stop-signal trials) on the SST (Verbruggen et al., 2019). After these deletions an adequately powered sample of 134 participants was retained, based on a set of Monte Carlo studies (i.e., computational experiments based on random sampling), which recommended 5–10 participants per parameter when conducting factor analysis (Hair et al., 2014; Kline, 2015). All continuous variables were normally distributed, with skewness and kurtosis values between −3 and 3 (Field, 2018). Therefore, multivariate normality was assumed, as the combination of univariate normally distributed variables often result in a multivariate normal distribution (Brereton, 2015).

Following the CFA, in order to examine the relationships between variables with fewer restrictions imposed, a principal axis factor analysis with direct oblimin rotation was performed using IBM® SPSS® Statistics Version 27. Direct oblimin rotation attempts to extrapolate the smallest number of factors with the highest correlations between variables and their respective factors whilst allowing for factors to correlate (Costello & Osborne, 2005). This form of rotation was selected as covariance between latent EF factors is commonly reported (Miyake et al., 2000). The KMO statistic verified the sampling adequacy for the analysis displaying a value beyond the acceptable limit of 0.5 (KMO = 0.673; Kaiser, 1974). Furthermore, Bartlett's test of sphericity was significant $\chi2(36) = 101.426$, $p < .001$, supporting the factorability of the correlation matrix.

## 3. Results

Preliminary analysis obtained the descriptive statistics and Pearson's correlation for variables assessed in the current study. Results are presented in Table 4.

### 3.1. Confirmatory factor analysis

The results of the CFA show that the three-factor yielded unsatisfactory levels of fit, $\chi2(26) = 49.918$, $p = .003$ only accounting for a total of 18.867 % of the variance, indicating that the predicted covariance matrices significantly deviated from the observed data. Conversely, the nested-factor models showed appropriate levels of fit, $\chi^2(24) = 36.360$, $p = .051$ accounting for a total of 30.067 % of the variance, meaning that the predicted covariance matrix did not significantly deviate from the observed data. Further goodness-of-fit-indices for each model are presented in Table 5.

As shown in Table 5 fit indices from both models failed to surpass the recommended cut of values. This indicates considerable discrepancy between the hypothesised models and the observed data (e.g., TLI < 0.90), and greater than acceptable standardised residuals between the hypothesised and observed covariance matrices (e.g., SRMR>0.08). The nested factor model did show one fit index within the recommended limit, as the RMSEA value was below 0.08. Therefore, when accounting for degrees of freedom, there were acceptable levels of discrepancy between hypothesised and observed covariance matrices. Figs. 3 and 4 present the path diagram for the three-factor and nested-factor models.

Factor loadings for the two models were mostly below 0.50. The three-factor model produced positive, significant loadings ($p < .05$) for the updating tasks. The shifting and inhibition factors failed to produce significant task loadings. The nested factor model weakened the loadings on the updating factor. The shifting factor still resulted in non-significant task loadings. The n-back, digit span, keep track, and verbal fluency were the only tasks to significantly load onto the common EF factor. Overall, the three-factor and nested-factor models showed poor levels of fit, and the shifting and inhibition tasks did not share enough variance to establish a latent factor.

**Table 4**
Descriptive statistics and correlations for EF measures used in the current study (*N* = 134).

| EF measure | Outcome measure | *M*(*SD*) | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. N-back | Average hits minus false alarms | 2.14(1.12) | – | 0.425** | 0.231** | −0.047 | 0.311** | 0.148 | 0.073 | 0.008 | 0.275** |
| 2. Digitspan backwards | Total correctly recalled trials | 6.98(2.05) | | – | 0.266** | 0.010 | 0.295** | 0.097 | 0.140 | −0.031 | 0.188* |
| 3.Keep track | Total correctly recalled words | 24.78(3.67) | | | – | −0.045 | 0.198* | 0.010 | −0.016 | 0.024 | 0.262** |
| 4. MCST | Total perseverative errors | 7.46(6.68) | | | | – | −0.015 | 0.132 | 0.045 | 0.072 | −0.147 |
| 5. Verbal fluency | Total unique words | 42.18(10.23) | | | | | – | 0.105 | 0.209* | −0.114 | 0.162 |
| 6. CST | Switch cost | 212.93(176.23) | | | | | | – | 0.172* | 0.051 | −0.116 |
| 7. SST | Integration method: SSRT | 177.28(68.03) | | | | | | | – | 0.063 | −0.030 |
| 8. Stroop | Trial 3: Trial 2 RT | 1.37(0.23) | | | | | | | | – | 0.041 |
| 9. CNT | Trial 3 + 4 efficiency score | 3.92(1.04) | | | | | | | | | – |

*Notes.* CNT = Contingency naming test; CST = Colour-shape task; EF = Executive function; *M* = Mean; MCST = Modified card sorting task; RT = Reaction time; *SD*=Standard deviation; SSRT = Stop-signal reaction time; SST = Stop-signal task.
[*] Significant at 0.05.
[**] Significant at 0.01.

**Table 5**
CFA fit indices for the three-factor and nested-factor models (*N* = 134).

| Fit Statistic | Three-factor Model | Nested Factor Model | Recommended Cut off Value |
|---|---|---|---|
| AGFI | 0.859 | 0.893 | ≥0.90 |
| SRMR | 0.103 | 0.086 | <0.08 |
| CFI | 0.651 | 0.819 | ≥0.90 |
| TLI | 0.516 | 0.729 | ≥ 0.95 |
| RMSEA | 0.083 | 0.062[a] | < 0.08 |

*Notes.* AGFI = Adjusted goodness of fit; SRMR = standardised root mean-squared residual; CFI = Comparative Fit Index; RMSEA = root mean square error of approximation; TLI = Tuker Lewis index.
[a] Within recommended cut off value based on Kline (2015).

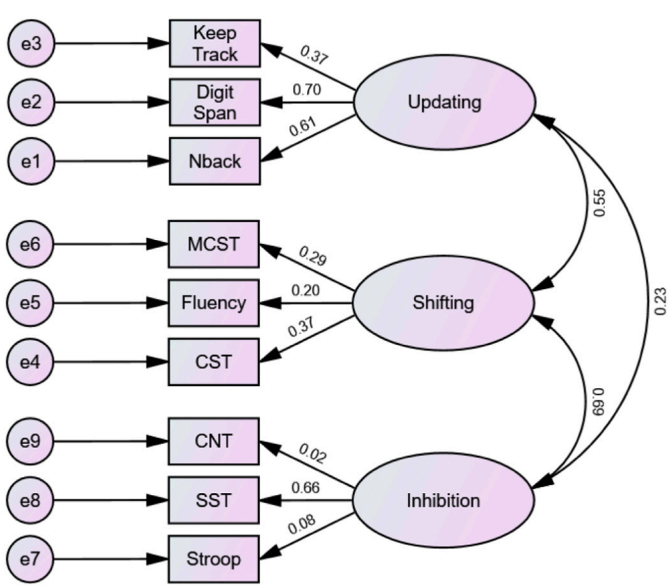


**Fig. 3.** Path Diagram from CFA of the three-factor model.
*Notes.* CNT = Contingency naming test; CST = Colour-Shape task; DB = Digitspan backwards; MCST = Modified card sorting task; SST = Signal-stop task.

### 3.2. Exploratory factor analysis

The CFAs were unable to demonstrate support for the three-factor or nested-factor model of the Unity/Diversity framework. Therefore, to continue the investigation of the underlying cognitive abilities being examined by the nine EF measures an EFA was conducted. The EFA produced a three-factor solution (eigen values > 1) accounting for a total of 50.852 % of the variance in the relationships amongst EF measures. Table 6 presents the factor loadings after varimax rotation. Factor 1 comprised of five measures digit span, n-back, keep track, CNT, and verbal fluency and accounted for 23.52 % of the variance. Factor 2 comprised of three measures MCST, CST, and SST and accounted for 15.37 % of the variance. Lastly, factor 3 comprised solely of the Stroop and accounted for 11.96 % of the variance.

Table 7 presents the correlation matrix for the identified factors following the EFA. Results showed negligible to small correlations between the three factors despite conducting an oblique rotation method.

## 4. Discussion

The current study aimed to re-evaluate the unity/diversity framework (Miyake et al., 2000) utilising EF tasks with sound psychometric properties and which purport to assess similar cognitive constructs, while remaining procedurally distinct. Despite being the most referenced conceptualisation of EF, the current study was unable to replicate either the unity/diversity three-factor or the nested-factor model, with both models displaying unsatisfactory levels of fit. Instead, using EFA the current study supported an alternative three-factor model containing an expanded working memory factor (Factor 1), a merged inhibition and shifting factor potentially reflecting cognitive flexibility (Factor 2), and an isolated Stroop factor (Factor 3).

The results of our CFAs conflicted with some previous research that successfully replicated the unity/diversity model in adolescents (Lee et al., 2013; Xu et al., 2013) and adults (Fournier-Vicente et al., 2008; Ito et al., 2015; Klauer et al., 2010). However, similar to previous research, the current study demonstrated strong factor loadings for the updating latent variable. Keep track has been frequently used in the previous literature, whereas digit span and traditional n-back have been less commonly used when constructing the updating factor. Regardless, updating measures are uniformly operationalised, with participants required to continuously update working memory with newly identified task relevant information, and scored based on the accuracy of their performance (Friedman et al., 2008; Jaeggi et al., 2010a; Redick & Lindsey, 2013). Therefore, the shared variance observed in performance on keep track, digit span, and the n-back supports a robust updating construct.

Where the current study diverges from the aforementioned studies is in the findings relating to shifting and inhibition, as no latent factors could be formed. Previous studies which uncovered a shifting factor, similar to Miyake et al. (2000), extensively relied on utilising three procedurally similar measures of task switching (Fleming et al., 2016; Fournier-Vicente et al., 2008; Friedman et al., 2011, 2016; Lee et al., 2013; Xu et al., 2013). For example, the colour-shape, picture-symbol, category switch, and number-letter tasks are all task switching paradigms which involve dichotomous responses where participants alternate between focusing on one of two stimuli characteristics (e.g., letter or number, colour or shape; Friedman et al., 2011, 2016). By exclusively using identical task switching paradigms with switch cost as their

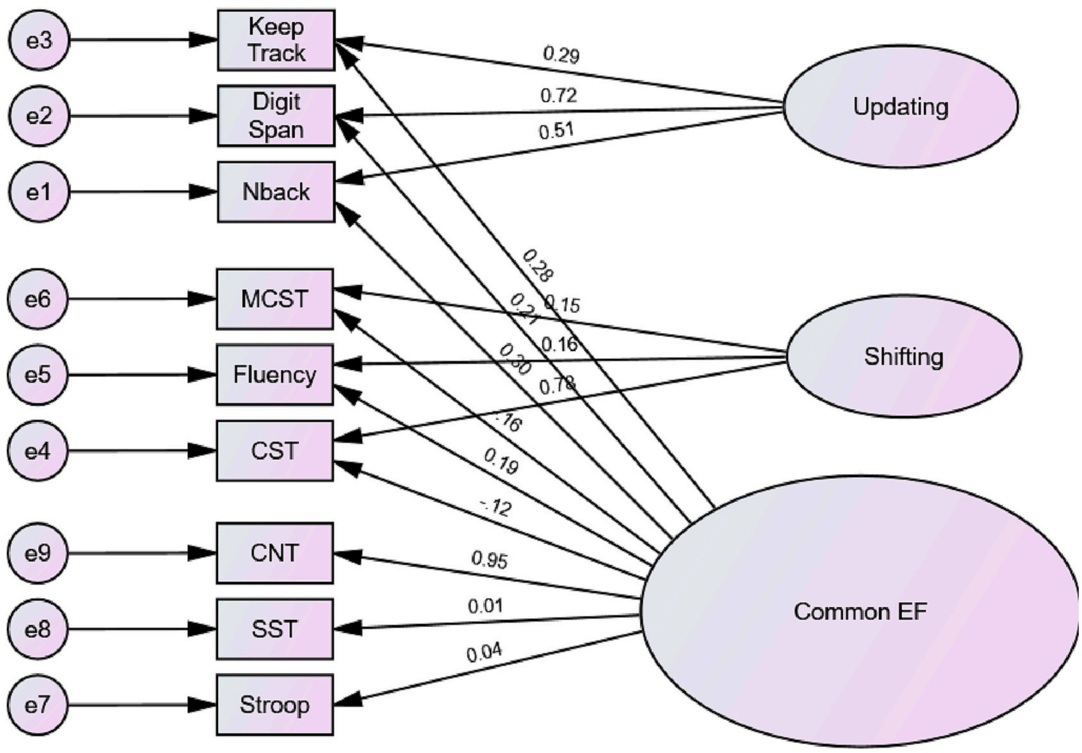


**Fig. 4.** Path Diagram from CFA of the nested-factor model.
*Notes.* CNT = Contingency naming test; CST = Colour-Shape task; DB = Digitspan backwards; MCST = Modified card sorting task; SST = Signal-stop task.

**Table 6**
Rotated pattern matrix loading of factors extracted by EFA ($N$ = 134).

| EF measure | Factor loadings | | |
|---|---|---|---|
| | Factor 1 | Factor 2 | Factor 3 |
| Nback | 0.705 | 0.155 | −0.041 |
| DB | 0.651 | 0.221 | −0.108 |
| Keep Track | 0.639 | −0.135 | 0.151 |
| Verbal Fluency | 0.487 | 0.267 | −0.401 |
| CNT | 0.650 | −0.379 | 0.154 |
| MCST | −0.147 | 0.513 | 0.282 |
| CST | 0.049 | 0.689 | 0.053 |
| SST | 0.098 | 0.589 | −0.099 |
| Stroop | 0.151 | 0.165 | 0.893 |

*Notes.* CNT = Contingency naming test; CST = Colour-Shape task; DB = Digitspan backwards; MCST = Modified card sorting task; SST = Signal-stop task.

**Table 7**
Component Correlation Matrix of Factors Extracted by EFA.

| Factor | 1 | 2 | 3 |
|---|---|---|---|
| 1 | – | 0.038 | −0.110 |
| 2 | | – | −0.058 |
| 3 | | | – |

common dependent variable, the shifting factor discovered in previous research may not be a true reflection of shifting attention between tasks and adjusting to rule changes, but instead may also reflect a specific procedural/fine motor skills (i.e., the ability to alternate keyboard presses in response to simplistic stimuli) or shared variance from an undesired cognitive ability central to task switching (e.g., processing speed, or inhibition from ignoring unnecessary task information; von Bastian et al., 2020; Yu et al., 2017).

The current study sought to compensate for these limitations and build a latent shifting factor by utilising the CST, used by Miyake et al. (2000), and two other procedurally distinct tasks that share a shifting element. The MCST is considered a measure of shifting as it requires participants to adapt to rule changes based on response feedback without being significantly influenced by participants' processing speed (Kopp et al., 2019b), and verbal fluency assesses participants ability to flexibility generate novel words and switch retrieval strategies to avoid repetitive ideas without the need to inhibit competing task rules (Aita et al., 2019; Paula et al., 2015). The inclusion of the MCST and verbal fluency in our shifting factor minimises the likelihood that the latent factor would represent other distinctive non-shifting cognitive abilities due to distinct task mechanisms. However, despite the theoretical similarities between the shifting measures selected, the current study was only capable of constructing a shifting factor with weak factor loadings. It is possible that the weak loadings were due to task impurity with shifting ability only contributing a small amount to performance on the three task and other unique task-specific factors accounting for the remaining unshared variance. Similarly, previous research has not always successfully replicated the shifting factor. When three identical task switching paradigms were not selected as measures of shifting, these studies favoured a two-factor model with shifting and inhibition merged into a single factor (Adrover-Roig et al., 2012; Bettcher et al., 2016; Brocki & Tillman, 2014; Brydges et al., 2012; Hull et al., 2008; Lehto et al., 2003; Monette et al., 2015; Usai et al., 2014). Therefore, the shifting factor constructed using task switching paradigms with overlapping procedures are more likely to be representative of task-specific variance, rather than shifting as a general cognitive construct.

The inhibition latent factor in the current study also failed to hold, however unlike the shifting factor which displayed weak factor loadings, the Stroop, SST, and CNT as shown in Table 4 shared negligible relationships with each other. Historically, the inhibition factor has always demonstrated the weakest factor loadings (Brydges et al., 2012; Hedden & Yoon, 2006; Hull et al., 2008; Rey-Mermet et al., 2019). Prior to the nested-factor model, studies supporting the three-factor model have successfully constructed an inhibition factor utilising the antisaccade task (Fleming et al., 2016; Friedman et al., 2011, 2016; Ito et al., 2015;

Klauer et al., 2010; Miyake et al., 2000) or flanker task (Bettcher et al., 2016; Lee et al., 2013). Critically, in recent investigations, these two tasks have shown stronger reliability properties than other inhibition measures (e.g., Stroop and SST; von Bastian et al., 2020). However, researchers have been questioning their construct validity as performance on the antisaccade and flanker tasks conflate inhibition control with other cognitive abilities (Hedge et al., 2018; Rey-Mermet et al., 2019; Rouder & Haaf, 2019). For example, unlike the Stroop or SST, the antisaccade and flanker tasks require the inhibition of environmental distractors rather than inhibiting a well learned habit (von Bastian et al., 2020). Therefore, it has been theorised that these tasks should primarily be considered as measures of attentional focus/sustained attention, with the inhibition component merely an incidental consequence of attentional fixation on the task goal (Hutton & Ettinger, 2006; White et al., 2011). Supportively, a cognitive control composite score constructed using the antisaccade and flanker task scores has been shown to significantly correlate with performance on the d2 test of attention ($r = 0.38$), which is widely considered a sustained attention task (Pahor et al., 2022).

The current study sought to address criticisms of inhibition measures outlined in recent publications (Draheim et al., 2021; Hedge et al., 2018; Rey-Mermet et al., 2019; Rouder & Haaf, 2019; von Bastian et al., 2020) by selecting task with demonstrated psychometric properties and are not subject to speed-accuracy trade-offs. The Victoria Stroop test was used as it has shown strong reliability properties (Strauss et al., 2006) and errors contribute to overall reaction time as participants are required to correct them. Additionally, the SST was used in the current study as it assesses an important inhibition mechanism through complete task interruption (von Bastian et al., 2020), however, the novel integration method of scoring (Verbruggen et al., 2019) was utilised over the mean estimation method used by Miyake et al. (2000). The integration method provides a more reliable assessment of participants' inhibition abilities as it is less biased by positive skews in their reaction time due to proportionally delayed responses (see Verbruggen et al., 2019). Lastly, the CNT was used as this task includes a dedicated inhibition score incorporating both participants' accuracy and speed, reflecting their capacity to supress previously learned rules to produce a less familiar response (Anderson et al., 2000; Riddle & Suhr, 2012). Despite these changes in task selection, the current study failed to construct a latent inhibition factor.

Currently, there is ongoing debate as to whether inhibition control is a unitary construct reflecting active maintenance of a goal whilst ignoring unnecessary information (Draheim et al., 2021; Hedge et al., 2018; von Bastian et al., 2020), or a multidimensional construct divided by task specific characteristics (e.g., mechanisms of the task, or type of distractors presented; Rey-Mermet et al., 2019; Rouder & Haaf, 2019; von Bastian et al., 2020). Indeed, the evidence from the current study lends credence to the argument that there are multiple task specific inhibition factors with unique learning processes, rather than a general inhibition control factor given the lack of relationships observed between the inhibition measures selected.

Miyake and colleagues supported the unitary view of inhibition control, eventually abandoning the inhibition factor in favour of a nested common EF factor (Friedman & Miyake, 2017). The common EF factor accounted for the shared variance between all nine EF measures and completely subsumed the inhibition factor. Miyake and Friedman (2012) theorised that inhibition measures are separated by task-specific mechanisms but are partially related due to a unitary cognitive construct that is represented in all EF tasks, the ability to control attention and bias cognitive operations towards a particular goal. In contrast, the current study did not support the nested-factor model which failed to achieve satisfactory levels of fit. These findings contradict previous research that showed support for this model (Fleming et al., 2016; Ito et al., 2015; Miyake & Friedman, 2012), however are in-line Karr et al.' (2018) systematic review that showed low rates of model acceptance for both the nested-factor and three-factor models across the literature. It is possible that satisfactory levels of fit for the common EF factor achieved in previous research was due to the use of reaction time as the scoring unit for most of the tasks selected (e.g., task-switching paradigms, Stroop, SST, go/no-go, and antisaccade). Reaction time scores are typically influenced by non-executive processes (e.g., general processing speed; Jurado & Rosselli, 2007), meaning that the common EF factor may have inadvertently reflected these processes (i.e., processing speed) rather than capacity to carry out goal-directed behaviour. The nested-factor model ultimately shares the same limitations outlined above for the three-factor model: (1) selecting shifting measures with overlapping task-switching paradigms, and (2) selecting inhibition measures with questionable psychometric properties. Despite attempts to address these issues in the current study, neither the three-factor model nor nested-factor model achieved satisfactory levels of fit.

The EFA in the current study successfully identified three uncorrelated factors. Whilst the model only accounted for half of the variance amongst the EF measures, it is important to emphasise that the complex nature of EF makes it difficult to isolate specific cognitive operations, which causes difficulties in measurement related to task impurity. EF tasks are often influenced by numerous lower-order cognitive abilities as well as other higher-order executive abilities, which can undermine relationships between measures (Hughes & Graham, 2002; Jurado & Rosselli, 2007; Miyake et al., 2000). Subsequently, the 50 % of variance accounted for by the EFA is considerable, especially considering the distinct procedures of the EF measures selected.

The first factor encompassed all three updating measures (i.e., keep track, DB, and n-back), along with verbal fluency and CNT, and likely represents a working memory factor. Working memory refers to a cognitive system involved in temporarily storing information for operation and manipulation (Baddeley & Hitch, 1974). Updating, while often used synonymously with working memory within EF literature, specifically refers to an individual's capacity to monitor and change the content of their working memory (Miyake et al., 2000; Packwood et al., 2011). It is unsurprising that the keep track, DB, and n-back loaded together, as this has been well supported in previous literature and all tasks involve rapidly updating the content of working memory to include new information from an ongoing list of stimuli (Hull et al., 2008; Ito et al., 2015; Klauer et al., 2010; Lee et al., 2013; Miyake et al., 2000; Xu et al., 2013). Successful performance on the Verbal fluency and CNT tasks is less dominated by the need to continuously update the contents of working memory, but nonetheless both tasks arguably have a strong working memory component as they require participants to mentally retain task rules while behaving in a goal directed manner. For example, during verbal fluency participants are required to mentally monitor the list of words they have provided against the list of task rules (e.g., don't use the same word more than once) in order to avoid errors (e.g., perseveration), at the same time as trying to retrieve further possible words. This is supported by previous research which showed superior working memory capacity was associated with improved performance on verbal fluency measures (Bittner & Crowe, 2007; Hedden & Yoon, 2006). Similarly, during the last two trials of the CNT participants are provided with multi-staged naming rules that they must mentally retain while completing the task. This likely engages considerable cognitive resources associated with their working memory capacity (Riddle & Suhr, 2012). Therefore, a working memory factor was identified with moderate to strong loadings from five out of nine tasks administered.

Factor 2 was determined to represent a broader cognitive flexibility construct, merging shifting tasks (i.e., CST and MCST) and inhibition tasks (i.e., SST and CNT). Shifting, as measured by task-switching paradigms (e.g., CST), only represents a singular component (i.e., subcategory) of the cognitive flexibility construct, being the ability to reallocate attention in response to environmental change. Cognitive flexibility refers to our ability to adapt thought patterns and behaviours in response to environmental changes (Diamond, 2013). Cognitive flexibility likely operates as a higher-order domain-general construct that is impossible to capture in a single score as the construct is foreseeably comprised of multiple lower-order task specific factors (see von Bastian et al., 2020 discussion on attentional control). All four tasks that loaded onto factor 2 appear to share two key task specific factors that are essential to adaptation in response to environmental change (i.e., cognitive flexibility). Firstly, attentional filtering, that is dealing with perceptual distractor interference through ignoring irrelevant/distracting information; and secondly, changing task representation, during certain trials participants were required to adjust their responses to accommodate changing task rules (von Bastian et al., 2020).

For the CST participants are required to filter out either the colour or shape of the stimuli presented based on an initial cue, and then break responding habits by adapting their response accordingly to either colour or shape (Miyake et al., 2000). In the MCST participants first isolate the card characteristic that aligns with the current sorting rule, ignoring other unnecessary characteristics, then must change their sorting rule after repeated successful trials (Nelson, 1976). During the SST, participants respond with either the left or right arrow key depending on the arrowhead presented, however when the stop signal is heard they must filter out the arrowhead and adjust their typical responses to no response (Verbruggen et al., 2019). Lastly, successful performance on the CNT requires filtering out unnecessary information (e.g., isolating the colour or shape of the object), and alternating between naming the shape or colour depending on the naming rule of each trial (Anderson et al., 2000). Notably, the CNT moderately loaded onto factor 2 despite strongly loading on factor 1. The CNT clearly contains a working memory component involved in mentally maintaining the multifaceted naming rule, however, these results show that cognitive flexibility is also instrumental to successful performance.

Lastly, factor 3 only contained a strong loading from the Stroop task and moderate loading from verbal fluency. The selection of the Victoria Stroop task was intended to alleviate the limitations of previous Stroop tasks (e.g., poor reliability and scores influenced by speed-accuracy trade-offs), which diminished its relationship with other measures (Guye & von Bastian, 2017; Rey-Mermet et al., 2019; Rouder & Haaf, 2019). Further, the Victoria Stroop did not load onto factor 2 despite the similar feature of filtering out distracting information based on habits (i.e., automatic reading response; von Bastian et al., 2020) shared with other tasks that loaded onto this factor (e.g., SST and CNT). Draheim et al. (2021) identified improved reliability characteristics and stronger relationships between the Stroop and other inhibition measures when accuracy differences rather than reaction time were used to score this task. Problematically, only 12 % of the sample in the current study performed an error on the Victoria Stroop, preventing accuracy from being used to construct inhibition scores.

It is unlikely that the Stroop represents an entirely distinct cognitive construct compared to the other tasks used in the current study as previous research has supported the Stroop as a measure of inhibition (Draheim et al., 2021; Gunten et al., 2020; Hedge et al., 2018). One possibility that may account for the current results is that the reaction time differences between the congruent and incongruent conditions of the Stroop, which are used to calculate the primary inhibition score, often display a small effect size (Rey-Mermet et al., 2019). Subsequently, Stroop inhibition scores may be unable to reliably rank-order individuals when accounting for other individual differences (e.g., processing speed) resulting in the sample variance being more indicative of error variance rather than differences in inhibition abilities. This may explain why Testa et al. (2012) who conducted an EFA on 19 EF measures showed that the Stroop loaded with completion time of trials 1 and 2 of the CNT, which are measures of general processing speed and not executive functioning (Anderson et al., 2000). This would also explain why verbal fluency moderately loaded onto this factor with the Stroop, as processing speed and the ability to verbalize responses quickly would contribute to successful performance for both tasks. Furthermore, both verbal fluency and the Stroop share a similar inhibitory component of supressing inappropriate verbal responses (Periáñez et al., 2021), in the case of the Stroop inhibiting the written colour word, whereas for verbal fluency inhibiting words that were already said.

Findings from previous lesion and neuroimaging studies shed further light on the two factors identified by the current study: working memory and cognitive flexibility. Whilst, to date no clear double dissociation has been demonstrated between EFs due to the highly integrative neural network of the frontal lobes (Fuster, 2015), some patterns are beginning to emerge. Firstly, compared to any other brain region the dorsolateral prefrontal cortex (DLPFC) is most associated with executive functioning (Fuster, 2015; Jones et al., 2012). Meta-analyses of lesion and neuroimaging studies have commonly implicated the large DLPFC in performance on EF measures (Alvarez & Emory, 2006; Ardila et al., 2018; Collette et al., 2006; Lee et al., 2017; Niendam et al., 2012; Nitschke et al., 2017). However, Stuss (2011) postulated that the dorsal regions appear to be more strongly associated with working memory, whereas the lateral regions are associated with cognitive flexibility.

The dorsal regions are notably involved in both spatial and verbal working memory (Crottaz-Herbette et al., 2004; Curtis, 2006; Narayanan et al., 2005; Ricciardi et al., 2006). Damage to the DLPFC has resulted in significantly worse performance on the n-back task for 3-back trials which require a higher degree of working memory, whereas performance on the 1- or 2-back trials – which have a lower cognitive load – was not significantly impaired (Barbey et al., 2013). Conversely, damage to lateral regions of the prefrontal cortex (PFC) resulted in significant impaired performance on a variety of cognitive flexibility measures (Cipolotti et al., 2016; Picton et al., 2007). Supportively, Tsuchida and Fellows (2013) compared 50 patients with PFC damage and 50 healthy controls and found damage to the lateral PFC regions was associated with impaired performance on both a Stroop task and task-switching paradigm. This finding highlighted that shifting and inhibition may rely on a similar neural network, and tentatively supports the current study's umbrella conceptualisation of cognitive flexibility. The potential double dissociation identified in these lesion and neuroimaging studies loosely supports the two factors identified in the current study. However, it is quite clear that successful executive functioning is dependent on an intact PFC as the various EFs operate in tandem to produce goal-directed behaviour (Fuster, 2015; Stuss, 2011), which supports the dual loadings of the CNT and verbal fluency.

The findings of both studies need to be interpreted considering some limitations. Sample size has often been criticised in studies conducting factor analysis. The current study recruited a sample comparable to other published studies investigating the factor structure of EF measures (Miyake et al., 2000; Testa et al., 2012) and in accordance with requirements outlines by factor analysis experts (Kline, 2015). However, it is important to note that a larger sample could have elucidated the relationships between measures more effectively (Karr et al., 2018). Beyond a larger sample, future researchers might also benefit from increasing the number of EF tasks administered to increase precision in the factors identified (Marsh et al., 1998), and to include higher-order cognitive abilities that have been overlooked (e.g., decision-making, strategic retrieval, planning; Karr et al., 2018) allowing for alternative conceptualisations of EF. Another potential limitation of the current study was that EF tests were administered over an online Zoom meeting due to COVID-19 pandemic restrictions, however the administration procedures for all measures were standardised across participants. It is unclear if online delivery may have impacted participant performance on the tasks. Regardless, it is important to note this procedural contrast with previous research. The unity/diversity framework has been consistently studied across all age groups, with one- and two-factor models showing better fit on child and older adult samples (Bettcher et al., 2016; Fournier-vicente et al., 2008; Hedden & Yoon, 2006; Ito et al., 2015; Lee et al., 2013; Wiebe et al., 2008, 2011; Xu et al., 2013). The findings of the current study exclusively focused on adults and obtained a primarily young adult sample. Further research is needed to re-evaluate the unity/diversity framework in younger and older samples. Additionally, middle-aged adults remain an understudied sample in EF research due to recruitment difficulties (Karr et al., 2018), and thus it remains unknown how or if the organisation of EF differs between younger and middle-aged adults, similar to the differences reported between younger and older adults (de Frias et al., 2009; Hedden & Yoon, 2006).

In conclusion, the findings of the current study failed to support the unity/diversity framework. Contrary to its intention, the unity/diversity framework is more representative of task-specific variance and does not seem to capture the theorised 'core' EFs. The three-factor and nested-factor models appear to only hold in specific conditions utilising carefully selected tasks with considerable procedural overlap. It might be tempting to attribute this finding solely to task impurity, however the results of the EFA provided contrary evidence. Over half of the variance was accounted for from nine procedurally distinct EF measures, with three factors identified: a working memory factor, cognitive flexibility factor, and a final factor which only encompassed the Stroop test. Crucially, our factors are not reflective of overlapping scoring methods or procedures, as both factors include a mixture of accuracy-based, time-based, and error-based scores. This further supports the notion that the identified factors tapped into an underlying cognitive construct. However, it is important to consider that some of the EF tasks (e.g., verbal fluency, and CNT) moderately loaded across two factors. This illustrates the difficulty of mapping a single cognitive ability to EF tasks as they are inherently designed to be complex to assess higher order cognitive functions, thereby resulting in task impurity (Hughes & Graham, 2002; Jurado & Rosselli, 2007).

Our EFA supported the validity of digit span, n-back, and keep track, while also demonstrating strong working memory components in verbal fluency and CNT. It is not surprising that working memory was found to be the most robust EF construct as it is amongst the most studied EF and is consistently well operationalised as it is arguably easier to isolate using statistical methods compared to other EFs (Diamond, 2013; Packwood et al., 2011). Crucially, the current study was unable to identify separate shifting or inhibition factors. Instead, the CST, MCST, SST, and CNT moderately-to-strongly loaded onto a single factor identified as a cognitive flexibility factor. Traditionally, cognitive flexibility has been viewed as an umbrella term that encompasses adaptive behaviour in pursuit of goal attainment, mentally maintaining multiple concepts simultaneously, and switching between different task rules (Diamond, 2013). Indeed, our study supports this conceptualisation of cognitive flexibility, however, also recognises there is considerable variance unaccounted for, presumably due to the distinct task mechanisms across measures of shifting or inhibition. Similar to the attentional control models outlined by von Bastian et al. (2020), cognitive flexibility may operate as a higher-order factor that can be divided into lower-order factors representing task specific variance due to unique learning processes stemming from different task mechanics (e.g., attention filtering, inhibition, self-monitoring etc.). Therefore, selecting a single test to represent cognitive flexibility may not be feasible, and future researchers and clinicians should carefully consider their task-specific skills of interest when deciding which EF tasks to administer. Overall, researchers need to continue attempts to refine existing measures and models of EF to ensure they are capturing complex, goal-directed behaviour as it occurs in the real world.

## Ethical approval

The study was approved by Victoria University Human Research Ethics Committee (reference number: HRE19-185).

## Funding

No grants or fellowships were used to fund this work.

## Declaration of competing interest

The authors declare that they have no conflict of interest in publishing this work.

## Data availability

The datasets generated and analyseNd during the current study are not publicly available due to ethical constraints but are available from the corresponding author on reasonable request and permission from the appropriate human research ethics committee.

## References

Adrover-Roig, D., Sesé, A., Barceló, F., & Palmer, A. (2012). A latent variable approach to executive control in healthy ageing. *Brain and Cognition, 78*(3), 284–299, 1016/j.bandc.2012.01.005.

Aita, S. L., Beach, J. D., Taylor, S. E., Borgogna, N. C., Harrell, M. N., & Hill, B. D. (2019). Executive, language, or both? An examination of the construct validity of verbal fluency measures. *Applied Neuropsychology: Adult, 26*(5), 441–451. https://doi.org/10.1080/23279095.2018.1439830

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*, 20–29, 1016/j.jecp.2009.11.003.

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review, 16*(1), 17–42. https://doi.org/10.1007/s11065-006-9002-x

Anderson, P. J., Anderson, V., Northam, E., & Taylor, H. G. (2000). Standardization of the contingency naming test (CNT) for school-aged children: a measure of reactive flexibility. In *, 1. Standardization of the Contingency Naming Test (CNT) for school-aged children: A measure of reactive flexibility* (pp. 247–273).

Ardila, A., Bernal, B., & Rosselli, M. (2018). Executive functions brain system: An activation likelihood estimation meta-analytic study. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 33*(4), 379–405. https://doi.org/10.1093/arclin/acx066

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–89). Academic Press.

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain & Education, 10*(1), 10–33. https://doi.org/10.1111/mbe.12100

Barbey, A. K., Koenigs, M., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human working memory. *Cortex; A Journal Devoted to the Study of the Nervous System and Behavior, 49*(5), 1195–1205. https://doi.org/10.1016/j.cortex.2012.05.022

von Bastian, C. C., Blais, C., Brewer, G. A., Gyurkovics, M., Hedge, C., Kałamała, P., & Wiemers, E. A. (2020, July 27). *Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations*. https://doi.org/10.31234/osf.io/x3b9k

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Bettcher, B. M., Mungas, D., Patel, N., Elofson, J., Dutt, S., Wynn, M., Watson, C. L., Stephens, M., Walsh, C. M., & Kramer, J. H. (2016). Neuroanatomical substrates of executive functions: Beyond prefrontal structures. *Neuropsychologia, 85*, 100–109. https://doi.org/10.1016/j.neuropsychologia.2016.03.001

Bittner, R. M., & Crowe, S. F. (2007). The relationship between working memory, processing speed, verbal comprehension and FAS performance following traumatic brain injury. *Brain Injury, 21*(7), 709–719. https://doi.org/10.1080/02699050701468917

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia, 5*, 135–140. https://doi.org/10.1016/0028-3932(67)90015-2

Brereton, R. G. (2015). The chi squared and multinormal distributions. *Journal of Chemometrics, 29*(1), 9–12. https://doi.org/10.1002/cem.2680

Brocki, K. C., & Tillman, C. (2014). Mental set shifting in childhood: The role of working memory and inhibitory control. *Infant and Child Development, 23*(6), 588–604. https://doi.org/10.1002/icd.1871

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

Channon, S. (1996). Executive dysfunction in depression: The Wisconsin card sorting test. *Journal of Affective Disorders, 39*(2), 107–114. https://doi.org/10.1016/0165-0327(96)00027-4

Cipolotti, L., Spanò, B., Healy, C., Tudor-Sfetea, C., Chan, E., White, M., Biondo, F., Duncan, J., Shallice, T., & Bozzali, M. (2016). Inhibition processes are dissociable and lateralized in human prefrontal cortex. *Neuropsychologia, 93*(Part A), 1–12. https://doi.org/10.1016/j.neuropsychologia.2016.09.018

Collette, F., Hogge, M., Salmon, E., & Van der Linden, M. (2006). Exploration of the neural substrates of executive functioning by functional neuroimaging. *Neuroscience, 139*(1), 209–221. https://doi.org/10.1016/j.neuroscience.2005.05.035

Costello, A., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*, 1–9.

Crottaz-Herbette, S., Anagnoson, R., & Menon, V. (2004). Modality effects in verbal working memory: Differential prefrontal and parietal responses to auditory and visual stimuli. *NeuroImage, 21*(1), 340–351. https://doi.org/10.1016/j.neuroimage.2003.09.019

Curtis, C. E. (2006). Prefrontal and parietal contributions to spatial working memory. *Neuroscience, 139*(1), 173–180. https://doi.org/10.1016/j.neuroscience.2005.04.070

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

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

Field, A. (2018). *Discovering statistics using SPSS* (5th ed.). SAGE Publications.

Fleming, K. A., Heintzelman, S. J., & Bartholow, B. D. (2016). Specifying associations between conscientiousness and executive functioning: Mental set shifting, not prepotent response inhibition or working memory updating. *Journal of Personality, 84*(3), 348–360. https://doi.org/10.1111/jopy.12163

Fournier-Vicente, S., Larigauderie, P., & Gaonac'h, D. (2008). More dissociations and interactions within central executive functioning: A comprehensive latent-variable analysis. *Acta Psychologica, 129*(1), 32–48.. https://doi.org/10.1016/j.actpsy.2008.04.004

de Frias, C. M., Dixon, R. A., & Strauss, E. (2009). Characterizing executive functioning in older special populations: From cognitively elite to cognitively impaired. *Neuropsychology, 23*(6), 778–791. https://doi.org/10.1037/a0016743

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

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

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology, 47*(5), 1410–1430. https://doi.org/10.1037/a0023750

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology, 52*(2), 326–340. https://doi.org/10.1037/dev0000075

Friedman, N. P., du Pont, A., Corley, R. P., & Hewitt, J. K. (2018). Longitudinal relations between depressive symptoms and executive functions from adolescence to early adulthood: A twin study. *Clinical Psychological Science, 6*(4), 543–560. https://doi.org/10.1177/2167702618766360

Frost, A., Moussaoui, S., Kaur, J., Aziz, S., Fukuda, K., & Niemeier, M. (2021). Is the n-back task a measure of unstructured working memory capacity? Towards understanding its connection to other working memory tasks. *Acta Psychologica, 219*, Article 103398. https://doi.org/10.1016/j.actpsy.2021.103398

Fuster, J. M. (2015). *The prefrontal cortex*. Academic Press.

Geurten, M., Vincent, E., Van der Linden, M., Coyette, F., & Meulemans, T. (2016). Working memory assessment: Construct validity of the Brown-Peterson test. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement, 48*(4), 328–336. https://doi.org/10.1037/cbs0000057

Goldstein, S., & Naglieri, J. A. (2014). *Handbook of executive functions*. Springer.

Gunten, C. D., Bartholow, B. D., & Martins, J. S. (2020). Inhibition tasks are not associated with a variety of behaviours in college students. *European Journal of Personality, 34*(3), 412–430. https://doi.org/10.1002/per.2250

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

Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis*. Pearson Education Limited.

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.supp

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

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions? *Child and Adolescent Mental Health, 7*(3), 131–142. https://doi.org/10.1111/1475-3588.00024

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

Hutton, S. B., & Ettinger, U. (2006). The antisaccade task as a research tool in psychopathology: A critical review. *Psychophysiology, 43*(3), 302–313. https://doi.org/10.1111/j.1469-8986.2006.00403.x

Isquith, P. K., Crawford, J. S., Espy, K. A., & Gioia, G. A. (2005). Assessment of executive function in preschool-aged children. *Mental Retardation and Developmental Disabilities Research Reviews, 11*(3), 209–215, 1002/mrdd.20075.

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology, 108*(2), 187–218. https://doi.org/10.1037/a0038557

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010b). The concurrent validity of the N-back task as a working memory measure. *Memory, 18*(4), 394–412. https://doi.org/10.1080/09658211003702171

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010a). The relationship between n-back performance and matrix reasoning — Implications for training and transfer. *Intelligence, 38*(6), 625–635. https://doi.org/10.1016/j.intell.2010.09.001

Jones, H. R., Jr., Srinivasan, J., Allam, G. J., & Baker, R. A. (2012). *Netter's neurology* (2nd ed.). Elsevier.

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review, 17*(3), 213–233. https://doi.org/10.1007/s11065-007-9040-z

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31–36. https://doi.org/10.1007/BF02291575

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, 144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

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. *Quarterly Journal of Experimental Psychology, 63*(3), 595–619. https://doi.org/10.1080/17470210903076826

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Publications.

Kopp, B., Lange, F., & Steinke, A. (2019a). The reliability of the Wisconsin card sorting test in clinical practice. *Assessment*, 1–16. https://doi.org/10.1177/1073191119866257

Kopp, B., Maldonado, N., Scheffels, J. F., Hendel, M., & Lange, F. (2019b). A meta-analysis of relationships between measures of Wisconsin card sorting and intelligence. *Brain Sciences, 9*(12). https://doi.org/10.3390/brainsci9120349

Lambek, R., & Shevlin, M. (2011). Working memory and response inhibition in children and adolescents: Age and organization issues. *Scandinavian Journal of Psychology, 52*(5), 427–432. https://doi.org/10.1111/j.1467-9450.2011.00899.x

Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A. M., & Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *British Journal of Educational Psychology, 82*(1), 82–99. https://doi.org/10.1111/j.2044-8279.2010.02016.x

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development, 84*(6), 1933–1953. https://doi.org/10.1111/cdev.12096

Lee, S.-H., Walker, Z. M., Hale, J. B., & Chen, S. H. A. (2017). Frontal-subcortical circuitry in social attachment and relationships: A cross-sectional fMRI ALE meta-analysis. *Behavioural Brain Research, 325*(Part B), 117–130. https://doi.org/10.1016/j.bbr.2017.02.032

Lehto, J. E., Juujarvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*(1), 59–80. https://doi.org/10.1348/026151003321164627

Lezak, M. D., Loring, D. W., & Howieson, D. B. (2004). *Neuropsychological assessment*. Oxford University Press.

Lineweaver, T. T., Bondi, M. W., Thomas, R. G., & Salmon, D. P. (1999). A normative study of Nelson's (1976) modified version of the WisconsinCard sorting test in healthy older adults. *Clinical Neuropsychologist, 13*(3), 328. https://doi.org/10.1076/clin.13.3.328.1745

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too Much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181–220. https://doi.org/10.1207/s15327906mbr3302_1

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. (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

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomason, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America, 108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Monette, S., Bigras, M., & Lafrenière, M. A. (2015). Structure of executive functions in typically developing kindergarteners. *Journal of Experimental Child Psychology, 140*, 120–139. https://doi.org/10.1016/j.jecp.2015.07.005

Narayanan, N. S., Prabhakaran, V., Bunge, S. A., Christoff, K., Fine, E. M., & Gabrieli, J. D. E. (2005). The role of the prefrontal cortex in the maintenance of verbal working memory: An event-related fMRI analysis. *Neuropsychology, 19*(2), 223–232. https://doi.org/10.1037/0894-4105.19.2.223

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12*(4), 313–324. https://doi.org/10.1016/S0010-9452(76)80035-4

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective & Behavioral Neuroscience, 12*(2), 241–268. https://doi.org/10.3758/s13415-011-0083-5

Nitschke, K., Köstering, L., Finkel, L., Weiller, C., & Kaller, C. P. (2017). A meta-analysis on the neural basis of planning: Activation likelihood estimation of functional brain imaging results in the tower of London task. *Human Brain Mapping, 38*(1), 396–413. https://doi.org/10.1002/hbm.23368

Packwood, S., Hodgetts, H. M., & Tremblay, S. (2011). A multiperspective approach to the conceptualization of executive functions. *Journal of Clinical and Experimental Neuropsychology, 33*, 456–470. https://doi.org/10.1080/13803395.2010.533157

Pahor, A., Mester, R. E., Carrillo, A. A., Ghil, E., Reimer, J. F., Jaeggi, S. M., & Seitz, A. R. (2022). UCancellation: A new mobile measure of selective attention and concentration. *Behavior Research Methods, 54*(5), 2602–2617. https://doi.org/10.3758/s13428-021-01765-5

Paula, J. J.d., Paiva, G. C.d. C., & Costa, D.d. S. (2015). Use of a modified version of the switching verbal fluency test for the assessment of cognitive flexibility. *Dementia & Neuropsychologia, 9*(3), 258–264. https://doi.org/10.1590/1980-57642015DN93000008

Periáñez, J. A., Lubrini, G., García-Gutiérrez, A., & Ríos-Lago, M. (2021). Construct validity of the stroop color-word test: Influence of speed of visual search, verbal fluency, working memory, cognitive flexibility, and conflict monitoring. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists, 36*(1), 99–111. https://doi.org/10.1093/arclin/acaa034

Perry, W., Potterat, E. G., & Braff, D. L. (2001). Self-monitoring enhances Wisconsin card sorting test performance in patients with schizophrenia: Performance is improved by simply asking patients to verbalize their sorting strategy. *Journal of the International Neuropsychological Society : JINS, 7*(3), 344–352. https://doi.org/10.1017/s1355617701733085

Picton, T. W., Stuss, D. T., Alexander, M. P., Shallice, T., Binns, M. A., & Gillingham, S. (2007). Effects of focal frontal lesions on response inhibition. *Cerebral Cortex (New York, N.Y.: 1991), 17*(4), 826–838. https://doi.org/10.1093/cercor/bhk031

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

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

Ricciardi, E., Bonino, D., Gentili, C., Sani, L., Pietrini, P., & Vecchi, T. (2006). Neural correlates of spatial working memory in humans: A functional magnetic resonance imaging study comparing visual and tactile processes. *Neuroscience, 139*(1), 339–349. https://doi.org/10.1016/j.neuroscience.2005.08.045

Riddle, T. L., & Suhr, J. A. (2012). Extension of the contingency naming test to adult assessment: Psychometric analysis in a college student sample. *The Clinical Neuropsychologist, 26*(4), 609–625. https://doi.org/10.1080/13854046.2012.666265

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

van Schooten, K. S., Duran, L., Visschedijk, M., Pijnappels, M., Lord, S. R., Richardson, J., & Delbaere, K. (2019). Catch the ruler: Concurrent validity and test-retest reliability of the ReacStick measures of reaction time and inhibitory executive function in older people. *Aging Clinical and Experimental Research, 31*(8), 1147–1154. https://doi.org/10.1007/s40520-018-1050-6

Sicard, V., Moore, R. D., Simard, A., Lavoie, G., & Ellemberg, D. (2022). Psychometric properties of a color-shape version of the switch task. *Applied Neuropsychology. Adult, 29*(5), 1020–1029. https://doi.org/10.1080/23279095.2020.1842410

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology, 6*, 1–24. https://doi.org/10.3389/fpsyg.2015.00328

Soveri, A., Lehtonen, M., Karlsson, L. C., Lukasik, K., Antfolk, J., & Laine, M. (2018). Test-retest reliability of five frequently used executive tasks in healthy adults. *Applied Neuropsychology. Adult, 25*(2), 155–165. https://doi.org/10.1080/23279095.2016.1263795

Strauss, E., Spreen, O., & Sherman, E. S. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press.

Stuss, D. T. (2011). Functions of the frontal lobes: Relation to executive functions. *Journal of the International Neuropsychological Society, 17*(5), 759–765. https://doi.org/10.1017/S1355617711000695

Testa, R., Bennett, P., & Ponsford, J. (2012). Factor analysis of nineteen executive function tests in a healthy adult population. *Archives of Clinical Neuropsychology, 27*(2), 213–224. https://doi.org/10.1093/arclin/acr112

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology, 14*(2), 167–177. https://doi.org/10.1016/S0887-6177(97)00095-4

Tsuchida, A., & Fellows, L. K. (2013). Are core component processes of executive function dissociable within the frontal lobes? Evidence from humans with focal prefrontal damage. *Cortex, 49*(7), 1790–1800. https://doi.org/10.1016/j.cortex.2012.10.014

Usai, M. C., Viterbori, P., Traverso, L., & De Franchis, V. (2014). Latent structure of executive function in five- and six-year-old children: A longitudinal study. *European Journal of Developmental Psychology, 11*(4), 447–462. https://doi.org/10.1080/17405629.2013.840578

Verbruggen, F., Aron, A. R., Band, G. P., Beste, C., Bissett, P. G., Brockett, A. T., Brown, J. W., Chamberlain, S. R., Chambers, C. D., Colonius, H., Colzato, L. S., Corneil, B. D., Coxon, J. P., Dupuis, A., Eagle, D. M., Garavan, H., Greenhouse, I., Heathcote, A., Huster, R. J.Boehler, C. N., … (2019). A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. *elife, 8*. https://doi.org/10.7554/eLife.46323

Wahlstrom, D., Weiss, L. G., & Saklofske, D. H. (2016). Practical issues in WISC-V administration and scoring. In L. G. Weiss, D. H. Saklofske, J. A. Holdnack, & A. Prifitera (Eds.), *Practical resources for the mental health professional, WISC-V assessment and interpretation* (pp. 25–62). Academic Press. https://doi.org/10.1016/B978-0-12-404697-9.00002-9.

Wechsler, D. (1997). *WAIS-III. Administration and scoring manual. The psychological corporation*. Harcourt Brace and Company.

White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology, 63*(4), 210–238. https://doi.org/10.1016/j.cogpsych.2011.08.001

Wiebe, S. A., & Karbach, J. (2018). *Executive function development across the life span*. Taylor and Francis.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*(2), 575–587. https://doi.org/10.1037/0012-1649.44.2.575

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology, 108*(3), 436–452. https://doi.org/10.1016/j.jecp.2010

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology, 33*(1), 101–111. https://doi.org/10.1080/13803395.2010.493149

Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PLoS ONE, 8*(10), 1–9. https://doi.org/10.1371/journal.pone.0077770

Yu, Q., Chan, C., Chau, B., & Fu, A. (2017). Motor skill experience modulates executive control for task switching. *Acta Psychologica, 180*, 88–97. https://doi.org/10.1016/j.actpsy.2017.08.013