A Framework for Executive Function in the Late Elementary Years

Paul T. Cirino[1], **Yusra Ahmed**[1], **Jeremy Miciak**[1], **W. Pat Taylor**[1], **Elyssa H. Gerst**[1], and **Marcia A. Barnes**[2]

[1]University of Houston

[2]University of Texas at Austin

Abstract

Objective—Executive Function (EF) is a commonly used but difficult to operationalize construct. In this study, we considered EF and related components as they are commonly presented in the neuropsychological literature, as well as those of developmental, educational, and cognitive psychology. These components have not previously been examined simultaneously, particularly with this level of comprehensiveness, and/or at this age range or with this sample size. We expected that the EF components would be separate but related, and that a bifactor model would best represent the data relative to alternative models.

Method—We assessed EF with 27 measures in a large sample (N = 846) of late elementary school aged children, many of whom were struggling in reading, and who were demographically diverse. We tested structural models of EF, from unitary models to methodological models, utilizing model-comparison factor analytic techniques. We examined both a common factor as well as a bifactor structure.

Results—Initial models showed strong overlap among several EF latent variables. The final model was a bifactor model with a common EF, and five specific EF factors (Working Memory-Span/Manipulation and Planning; Working Memory-Updating; Generative Fluency, Self-Regulated Learning; Metacognition).

Conclusions—Results speak to the commonality and potential separability of EF. These results are discussed in light of prevailing models of EF, and how EF might be used for structure/description, prediction, and for identifying its mechanism for relevant outcomes.

Keywords

executive function; framework; factor analysis

Executive function (EF) is a domain-general control process important for managing goal-directed behavior. However, any straightforward definition belies the variability in the way EF is conceptualized within and across literatures. One approach addresses functions that are disrupted with damage to the frontal lobes (e.g., Stuss, 2011), and includes group or

individual studies of brain injury (e.g., Tranel, Anderson, & Benton, 1994; Welsh, Pennington, & Grossier, 1991), and/or structural and/or functional imaging to evaluate these associations (e.g., Alvarez & Emory, 2006; Jonides et al., 1993). A second approach delineates a collection of separate but related cognitive functions (e.g., shifting, inhibition, working memory, planning, fluency), through theory (e.g., working memory and inhibition, Roberts & Pennington, 1996) or empirical study (e.g., factor analysis; e.g., Miyake et al., 2000; Wiebe, Espy, & Charak, 2008), that have in common the management or facilitation of goal-directed behaviors. A third approach identifies problem solving stages of representation and preparation, generating and implementing solutions, performance monitoring, and reflection. These stages have been considered from the perspective of development (e.g., Zelazo, Carter, Reznick, & Frye, 1997), from neuropsychology and rehabilitation (e.g., Levine et al., 2000; 2011; Ylvisaker & Feeney, 2002), or from education (e.g., Zimmerman, 1989; 2000; Pintrich, 2000; 2004). Terms used to reflect this process include self-regulation (and in particular, self-regulated learning), metacognition, executive/attentional control, executive attention, and executive function, among others.

The goal of the present study was to provide a framework for EF, mainly relying on the second and third approaches noted above, and utilizing factor analytic methodology. While imaging or lesion data was not emphasized, the measurement structure identified could provide impetus for studies that do. We considered EF to be an umbrella term with both cognitive and self-regulatory aspects (Hofmann, Schmeichel, & Baddeley, 2012; Ilkowska & Engle, 2010).

We evaluated students in grades 3 through 5 (ages 8 to 11) for three interrelated reasons. First, developmentally sensitive change occurs cognitively and biologically across this age span, with the development of EF following a nonlinear trajectory corresponding with the physical maturity of the frontal lobes (Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Best, Miller, & Jones, 2009; Garon, Bryson, & Smith, 2008; Jurado & Roselli, 2007; Luciana & Nelson, 2002; Yakovlev & Lecours, 1967). Second, most existing structural studies of EF focus on preschool or adult age ranges; in contrast, a key question the present study is well equipped to address is whether EF components reflect a more unitary versus segregated group of processes at this intermediate age range. Third, students in this age range are expected to have "mastered" basic academic skills, and greater independence, initiative, and planning is required as curricula shift towards developing and applying generalized learning skills to new contexts (Speece, Ritchey, Silverman, Schatschneider, Walker, & Andrusik, 2010). EF becomes a key support as students integrate and regulate academic learning and understanding within and across domains.

The present study emphasizes a large and diverse urban public school sample. Many of these participants are of minority status, are raised in lower SES households, and some may speak languages in addition to English, which is representative of most large urban school districts (National Center for Education Statistics, 2016). These characteristics are associated with lower levels of achievement and/or EF (Hackman & Farah, 2009; Hackman, Gallop, Evans, & Farah, 2015; Raver, Blair, Willoughby, & The Family Life Project Key Investigators, 2013). We oversampled students in grade 4 with reading difficulty because this work overlapped with a large-scale intervention study (though the data reported preceded rather

than followed intervention; see Methods below). Doing so was necessary to improve feasibility (and efficiency) for both studies, but evaluating the structure of EF in this context is also relevant to a population where supports for EF are likely to be appropriate.

Executive Function Components

We identified eight EF components across neuropsychological, cognitive, developmental, and educational literatures: (1) working memory; (2) inhibition; (3) shifting; (4) planning; (5) generative fluency; (6) self-regulated learning; (7) metacognition; and (8) behavioral regulation. The first three of these were encapsulated in the work of Miyake et al. (2000), but the other five are distributed across the literatures noted above. No single work has considered all of the above simultaneously. It is possible to argue for the inclusion versus exclusion of any of these individual components, or to separate them by complexity, or even as "outcomes" of EF. We took an inclusionary approach because different literatures use different terminology and different measures. If each of the domains examined here are equally representative of EF, than they should relate to one another to some degree, but also share commonality. To the extent that a particular EF component "does not fit" then it should separate from the others, and the results of this study will be well-positioned to understand whether and along what lines such distinctions might occur.

Baddeley, Allen and Hitch (2011) describe the overlap of *working memory (WM)* with EF. Theorists such as Cowan (2008) and Engle (2002) emphasize WM as the active portion of long-term memory, such that it can be proffered as a substitution for EF itself – that is, EF theories become theories about how working memory operates. For this study, we chose working memory tasks that required: (a) recall with manipulation; or (b) simultaneous processing and storage; or (c) ongoing maintenance/updating of information in active memory. *Inhibition* is also featured in many studies of EF, though inhibition can be difficult to separate from a common EF factor (e.g., Miyake & Friedman, 2012). A key requirement is that a prepotent response is inhibited, and this was featured in the measurement of this domain for this study. *Shifting* as considered here requires a move from one cognitive set of rules, sequences, or patterns, to another, as similarly defined in other studies (Miyake et al., 2000). *Planning* has often been assessed as a clinical task (Delis, Kaplan, & Kramer, 2001), and may be considered a "higher-level" EF (Miyake et al., 2000). Solving a problem within a set of rules/parameters most closely resembles the process function of EF because strategies must be selected, implemented, and monitored. Tests of *Generative Fluency* have a lengthy history in neuropsychology (Milner & Petrides, 1984), with the key component being efficient and accurate retrieval from semantic memory within a set of parameters.

Models of *Self-Regulated Learning (SRL)* (e.g., Pintrich, 2000; 2004; Zimmerman & Schunk, 2011) focus on how individuals control their own learning processes. Zimmerman's (2000) model includes phases of forethought, performance, and self-reflection; Pintrich (2004) emphasized planning, monitoring, control, and reflection. These phases are similar to the frameworks of EF described by Zelazo et al. (1997) and Ylvisaker & Feeney (2002). SRL is multifaceted, but here we focused on strategy use and self-efficacy, as well as perceived skill and preference for reading (as a proxy for motivational stance), as these imply knowledge of one's own abilities. An important methodological difference in

comparing the literatures of SRL with EF is that SRL is often assessed via self-report, whereas EF is usually assessed via cognitive performance measures (particularly the "cool" variety; e.g., Welsh & Peterson, 2014).

Metacognition and EF are both often described as superordinate processes – those that monitor, manipulate, or control/regulate other cognitive processes, and metacognitive processes are featured in some models alongside EF (e.g., Stuss, 2011). Metacognition has also been perceived as a "bridge" function, for example, between EF and motivation (Borkowski, Chan, & Muthukrishna, 2000); Barkley's model (1997; 2014) considered metacognition as a component of EF (reconstitution). The Behavioral Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, Kenworthy, 2000) considers Metacognition as one of its primary scales.

In contrast to SRL, *Behavioral Regulation* models are often concerned with the ability of individuals to control their behavioral actions and/or emotions (Blair, 2002; Barkley, 1997; 2014; Posner & Rothbart, 2000; Raver, 2004). In some ways, this is similar to the idea of "hot" executive functions, where the situation is in some way emotionally valanced (Allan & Lonigan, 2011; Zelazo & Carlson, 2012; Welsh & Peterson, 2014). Although decision-based cognitive tasks are available, most often behavioral regulation is assessed through self- or other-ratings, especially beyond the preschool age. Few studies evaluate the relationship of neuropsychological measures of EF to behavioral self-regulation measures beyond the earliest elementary ages, though where examined, relationships are generally small (e.g., Duckworth & Kern, 2011).

Prior Factor-Analytic Studies in School-Aged Children

The literature on EF is vast, and many factor analytic studies have been conducted. However, none have matched the scope (e.g., number of measures, across different literature perspectives) and sample size evaluated here, particularly in this sensitive age range (middle to late elementary students). Several recent studies have focused on preschool children using careful measurement approaches, identifying a unitary EF factor (Wiebe et al., 2008; Willoughby, Blair, Wirth, & Greenberg, 2010). On the other hand, latent variable studies in adults show that EF components are notably fractionated though not independent (e.g., Miyake, 2000; Hull, Martin, Beier, Lane & Hamilton, 2008; Kane, et al., 2004). Structural results in school-aged samples vary, often finding factors associated with WM or updating, inhibition, and shifting/flexibility, though not always together. For example, sometimes a combination of shifting and inhibition is included, as a latent factor (Brydges, Fox, Reid, & Anderson, 2014; Lee, Bull, and Ho, 2013). Only a few studies have considered factors other than working memory, inhibition or shifting (Floyd, Bergeron, Hamilton, & Parra; 2010; Gioia, Isquith, Retzlaff, & Espy, 2002), making it unclear how additional domains such as fluency, planning, or self-regulated learning may or may not fit. The study by Huizinga et al. (2006) utilized a sophisticated modeling approach and covered the age range from 7 to 21, with a relatively large sample size (~400). All tasks were computerized and involved speeded choice reaction times. Controlling for processing speed, factors of working memory and shifting were obtained, though not inhibition. Most prior studies captured only performance-based (e.g., paper-and-pencil, computerized, and/or experimental) measures of

EF – none included self-report measures, with the exception of Gioia et al. (2002) where the structure of EF was examined using the BRIEF.

EF Factor Models: Common versus Bifactor?

There is debate about how EF components aggregate with or disaggregate from one another. For example, Willoughby, Pek, and Blair (2013) noted the typically weak interrelationships of factor indicators of EF. Factor analytic choices typically include a common factor model, a hierarchical factor model, and a bifactor model (Canivez, 2014). Common factor models comprise the vast majority of prior work, specifying which latent factors cause which indicators. Hierarchical factor models posit factors that have direct effects on indicators, but go on to posit second-order factors that have only indirect effects on indicators (through the specific factors); the second order factor also is presumed to account for all the relationships of the specific factors with one another. Bifactor models also posit general and specific factors, but in these models, indicators have two direct effects – one from the general factor, and another from the specific factor to which it is assigned. Specific factors are typically uncorrelated with both the general factor, *and* with one another; correlations among measures from the same latent variable arise from the general and specific factors, whereas correlations among measures from different factors arise from the general (EF) factor only (Holzinger & Swineford, 1937; Reise, 2012). In this study, we compare the common factor model to the bifactor case, given its consistency with the Miyake and Friedman models (Friedman, Miyake, Robinson, & Hewitt, 2011; Friedman et al., 2008; Miyake et al., 2000) in adults, and with a more general conceptualization of EF as "separate but related" function(s).

Summary and Hypotheses

The present study extends prior work in at least three ways. First, we address the need for a clearer understanding of the structure of EF in the school-age range. Second, measures from different literatures are simultaneously evaluated. Third, the large sample in a relevant context is novel, and we employed advanced modeling techniques to examine EF comprehensively. An overarching goal was to align the way that EF is conceptualized across literatures while also reflecting the myriad ways it is operationalized.

Our primary hypothesis was that we could model together numerous observed indicators at the confirmatory factor analytic (CFA) and bifactor level. To the extent that different literatures describe similar conceptualizations of domain-general control processes important for managing goal-directed behavior (sometimes calling it EF, other times working memory or executive attention, other times self-regulation), then measurement of each should bear relation to one another. However, to the extent that these measurements differ either methodologically (e.g., ratings versus laboratory tasks), or as lower versus higher level (e.g., "predictors" versus "outcomes"), then latent factors of these should be clearly separable from the others. These issues reflect a key debate regarding EF – whether it is unitary versus fractionated. Because both are likely true to some degree true, a bifactor model should fit the data better than a common factor model (8 correlated domains), and better than a single-factor model. We also specifically compared unitary, common, and

bifactor models against those driven by methodological factors (e.g., type of measure – rating scales or performance measures; type of content – visual or verbal); these should fit the data more poorly than our hypothesized bifactor model.

Second, we expected predictable relationships with demographic factors including age, sex, economic disadvantage, sex, and English proficiency status. We expected that any common EF factor (resulting from a bifactor model) would relate positively with age (older students having stronger performances). Age was examined rather than grade, as EF domains are not typically part of school curricula. We did not have a priori expectations for sex differences in EF. We expected that a common EF factor would correlate negatively with economic disadvantage and English proficiency (disadvantaged students or those with of limited English proficiency having weaker EF), which may in part be due to the language demands of many of the tasks.

Methods

Participants

Participants were 846 students in grades 3 to 5 (age range 8 to 12 years). Table S1 (see Supplementary Files) presents demographic characteristics (N, sex, age, site, ethnicity, economic disadvantage, special education status, and limited English proficiency status), by grade and for the overall sample. The sample was diverse ethnically, and a majority were economically disadvantaged. Students were drawn from 147 classrooms 19 schools. However, we chose not to fit multilevel models because the study was not designed to address multilevel questions about between-classroom level interrelations of EF factors. EF is not instructed the way core academic skills are, making it ambiguous to hypothesize about classroom effects.

Measures

Descriptive information for all measures by EF factor is provided in Table 1. For space considerations, additional information about task requirements, administration details, and psychometric properties can be found in Supplement 2. Raw scores instead of norm-referenced standard scores were utilized in analyses for consistency and because relations with age were examined directly.

Working Memory was assessed with: (a) Working Memory Test Battery for Children (WMTB-C; Pickering & Gathercole, 2001) Listening Recall subtest; (b) Inquisit Corsi Blocks (Inquisit 3, 2003) was included as a measure of span; and (c-f) four n-back measures were used, from the same platform (Inquisit 3, 2003); two that utilized letters versus complex (un-nameable) shapes, and each had a 1-back and 2-back version. *Inhibition* used: (a) Inquisit Cued Go-No-Go task (Inquisit 3, 2003); and (b) Stop Signal task (Inquisit 3, 2003). *Shifting* used specific components of four tests of the Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2001) that required alternation: (a) Design Fluency; (b) Verbal Fluency; (c) Color-Word Identification Test; and (d) Trail Making Test. *Planning* was assessed with: (a) Tower of London (Inquisit 3, 2003); and Woodcock-Johnson-III Planning subtest (Woodcock, McGrew & Mather, 2001). *Generative Fluency* used three D-

KEFS (Delis et al., 2001) fluency measures (Letter Fluency, Category Fluency, Design Fluency) that did not require alternation. *Self-Regulated Learning* was assessed with three subscales (self-efficacy and effort, strategies, and perceived skill and preference for reading) derived from a prior factor analytic study of the Contextual Learning Scale (Author, 2012), a self-report measure derived from previously used and novel items. *Metacognition* was assessed with teacher-reported subscales (Initiate, Monitor, Planning/Organization, and Working Memory) of the Metacognition Index from the Behavioral Rating Inventory of Executive Function (BRIEF; Gioia et al., 2000), and with teacher ratings of the Inattention subscale of the Strengths as Weaknesses of Attention-deficit hyperactivity disorder symptoms and Normal behavior (SWAN, Swanson et al., 2006). *Behavioral Regulation* was also assessed with the BRIEF (Gioia et al., 2000) and the SWAN (Swanson et al., 2012), using alternate subscales (Inhibit, Shift, and Emotional Control from the BRIEF, and Hyperactivity/Impulsivity from the SWAN).

Planned Missing Data

The current study reports on 27 different measures, and it was not feasible to administer all measures to all students. Rather than restrict the number of measures (and reduce construct coverage), we opted to decrease some covariance coverage (the number of students that had data for any pair of measures to establish their relationships). Missing data are problematic if they are not missing completely at random (MCAR) or missing at random (MAR) (Rhemtulla & Little, 2012; Rubin, 1976). We empirically tested the tenability of the MCAR assumption using Little's MCAR test (Little, 1988), and the results suggested that the data were not MCAR (χ^2 (1927) = 2105.27, χ^2/df = 1.09 p = 0.03). However, MCAR is a stringent assumption which may not be tenable in practice (Muthén, Kaplan & Hollis, 1987; Rhemtulla & Hancock, 2016). Modern methods for handling missing data, such as full information maximum likelihood (FIML), allow MCAR assumptions to be relaxed to allow for data to be MAR such that the likelihood of missingness on a given variable is conditional on other variables in the data (Enders & Bandalos, 2001). The primary issue is power, which is ameliorated with FIML (Little, Jorgensen, Lang, & Moore, 2013), so that all observations, measures, and available data points are utilized. Therefore, we followed these techniques in the present study. The specific method we used was a variant of the multiform approach (Rhemtulla & Little, 2012). Students were randomly assigned within grade to receive a "pattern" of tests. In practical terms, 9 of the 19 different tests were included in all patterns (though some missingness within pattern was also randomly assigned for three measures). There were six patterns (see Table 1) that produced covariance coverage of 15% to 98% for any pair of variables, and the proportion of students who received each pattern was similar, (e.g., overall and within grade p = .981 and p = .990, respectively). A proportion of students in grade 4 were designated as limited English proficiency, though each pattern also included similar proportions with and without this status, p = .938.

Procedures

Performance data were collected in schools by well-trained examiners with on-site supervision, primarily in Fall; teacher-rated measures were collected in late Spring. All performance based measures were collected individually; group-administration was used for student rating scales and for the screening measure of reading. Total assessment time was

approximately 4 hours, distributed across 4 or 5 sessions typically distributed across a week. Test orders within the randomly assigned patterns were not fixed. All research procedures were approved by participating Institutional Review Boards.

Analytic Approach

Variable distributions and outliers were specifically evaluated. Where issues with test administration or scoring were flagged, source documents and observations were examined for serious concerns regarding data appropriateness. This resulted in 84 data points being trimmed (set to missing). Also, if performance was well outside typical bounds (+/− 3 SD and ~ 0.5 SD from the next nearest score) then data were Winsorized (placed just beyond the next nearest value and so maintaining rank ordering) similar to criteria used elsewhere (Willcutt, Pennington, Olson, Chhabildas, & Hulslander, 2005); this was true for 46 additional data points. It should be noted that these 130 data points comprised only 0.83% of the total number of points available for analyses. The Stop Signal Test was an exception; here, 161 data points were excluded due to aberrant performance; cutoffs (Schachar, Mota, Logan, Tannock, & Klim, 2000) were less than 67% correct on go trials (n=116), and/or less than 12.5% correctly inhibited stop trials (n = 21); 24 students met both cutoffs. There were no differences between students whose Stop Signal score was versus was not included, in terms of pattern received, grade, age, sex, race, economic disadvantage, or limited English proficiency status (all p > .05), although those who were excluded had lower vocabulary scores than those whose scores remained, p = .008, d = .24).

For all analyses, a "weight" was assigned to each observation. Because of oversampling for reading difficulty in grade 4, reading scores in this grade (and presumably, correlates of reading skill) would be systematically lower than expectation. Weighting applied during analyses has the effect of decreasing the impact of the oversampled group on the overall mean, which in turn becomes more representative of the population from which they were drawn (in our case, the 19 schools across three school districts with whom we worked). Weighting has its largest effects on means, which was less relevant for the purposes of the present study, but it can also impact covariances. However, it should be noted that we also ran our final models with and without weighting and results were highly similar. Specific details regarding weighting appear in Supplement 3.

The primary analytic technique was factor analyses. Our proposed framework was designed to consider the eight domains in a confirmatory factor analytic (CFA) measurement model as well as its counterpart bifactor model. We ran models in MPlus v 7.1 (Muthén & Muthén, 2012) with robust full-information maximum-likelihood estimation to establish model fit. These models revealed several obvious sources of misfit, the details of which are presented in Supplement 3; for brevity, the final model is emphasized below. We evaluated the fit of all models at both the global and local level. For global fit, we used common measures, and accepted cut-offs (Hu & Bentler, 1999; Kline, 2011): Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) > .90 to .95; χ^2; root mean squared error of approximation (RMSEA < .06), and squared root mean residual (SRMR < .08). For local fit, we examined parameters such as loadings, standard errors and residuals. In all models, indicators were scaled so that higher scores were associated with better performance. For comparison to

methodological models and other model comparisons, we used χ^2 differences, and changes in global measures such as Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and Adjusted BIC (of ~ 10; Burnham & Anderson, 2004). Once all models has been evaluated, we correlated age, sex, economic disadvantage, and English proficiency status with the factors to evaluate their relations.

Results

Table 1 presents the means, standard deviation, skewness and kurtosis values for each measure. Table S2 presents the correlations among measures.

Common Factor Models

Model specifications and fit statistics for four common factor models are reported in Table 2. The first model (Model 1a) evaluated the eight proposed latent variables, and was a poor fit to the data due to latent variables that were too strongly correlated. The revised model (Model 1b) fit the data much better, but the Working Memory (WM) correlated strongly with several others, and factor loadings suggested a distinction among measures of span and manipulation, versus updating. This is in line with recent debate about the extent to which these two types of tasks are more similar versus different (Schmiedek, Hildebrandt, Lovden, Wilhelm, & Lindenberger, 2009; Redick & Lindsey, 2013; Schmiedek, Lovden, & Lindenberger, 2014). Therefore, we separated the original WM factor into separable components of Span/Manipulation (WM-S/M) versus Update (WM-U). This model (Model 2a) did not converge, as the span and manipulation WM factor was strongly correlated with Shifting, and so were combined, resulting in a final common factor model (Model 2b) with seven factors: (1) WM-S/M with Planning; (2) WM-U; (3) Inhibition; (4) Shifting; (5) Generative Fluency; (6) Self-Regulated Learning; and (7) Metacognition (MCOG). Further details regarding model revisions are in Supplement 3. Table 1 provides maximal reliabilities (ρ^*, Raykov & Hancock, 2005) from Model 2b; as expected, factor reliabilities were better than the reliability of any separate indicator. It was also the case that factors with more reliable indicators had higher maximal reliability; the only noticeably low value was for Inhibition ($\rho_{inhibition}= 0.54$). Table 3 contains common factor intercorrelations.

With Model 2b finalized, we compared this to alternative models, including a bifactor analogue. A unitary model (Model 3a) was considered, one that is nested within Model 2b (using the same 27 indicators of Table 1). The unitary model was a substantially worse fit relative to Model 2b. A variety of methodological models were also evaluated and each also produced a significantly worse fit to the data than Model 2b (see Table 2).

Bifactor Model

We had hypothesized that a bifactor model (Model 4) would fit the data better than its common factor analogue (Model 2b). The initial bifactor model did not converge because the Shifting and Inhibition specific factors did not identify their respective indicators (i.e., specific loadings were very low) in the context of a common EF factor. Therefore, these two specific factors were dropped leaving their indicators to load only on common EF. The final bifactor model (see Figure 1) had five specific factors (WM-S/M and Planning, WM-U,

Generative Fluency, Self-Regulated Learning, and MCOG), and was clearly a better fit to the data than Model 2b by most criteria (e.g., χ^2 difference, change in CFI, AIC, and BIC).

Table 4 shows the pattern of loadings for common EF and specific factors of Model 4. Loadings were stronger for the indicators of three specific factors (WM-U, Self-Regulated Learning, and MCOG) relative to the common EF factor. Conversely, loadings were higher for the specific WM-S/M and Planning factor. Fluency factor loadings were intermediate but generally larger for common EF. Indicators of Shifting and Inhibition loaded only on the common EF factor, though these were larger and more consistent for Shifting relative to Inhibition. Maximal reliabilities were also computed for the specific factors and the general EF factor from Model 4. From Model 2b to Model 4, reliability was highest for the EF factor (ρ_{EF}=0.97), and was also high for three of the specific factors from Model 4 (ρ_{WMU}= 0.82, $\rho_{self\_regulation}$=0.84, and $\rho_{meta\_cognition}$= 0.97). Reliability was moderate for two specific factors ($\rho_{WMSM}\rho_{WM/planning}$ = 0.66, and $\rho_{fluency}$= 0.79). The average intercorrelation among all 27 indicator variables (see Table S2) was $r = .17$ (median $r = .15$), though the intercorrelations within each of the specific factors that remained in the bifactor model were appreciably higher (WM/SM and Planning, median $r = .28$; WMU = .41; SRL = .52; Generative Fluency = .32; MCOG = .77).

Individual and Group Differences

With the finalized structural model, we related the resultant general and specific EF factors to demographic factors of age, economic disadvantage, sex, and language proficiency status. The model converged but the robust chi-square was not computable, though SRMR = .063, and parameter estimates were provided, which were of primary interest.

Age was positively correlated with general EF, $r = .38$, $p < 0.001$, and negatively correlated with specific Metacognition, $r = -.19$, $p < 0.001$ factors. All other age correlations were smaller in magnitude (range $r = |.03$ to $.09|$) and non-significant. Economic disadvantage correlated negatively with EF, $r = -.20$, $p = 0.003$, such that receipt of free lunch was associated with lower EF performances; there were no significant relations with other factors. Sex was positively correlated with WM-U, $r = .15$, $p = 0.017$; with SRL, $r = .14$, $p = 0.001$; and with Metacognition, $r = .27$, $p < 0.001$. These correlations were modest and conferred a slight advantage for girls; for WM-S/M and Plan, and for Fluency, advantages for boys were of similar magnitude ($r = -.13$ and $r = -.14$, respectively), though neither were significant. Finally, for students designated as having limited English proficiency, their performances were lower on EF, $r = -.26$, $p < 0.001$, and Fluency, $r = -.25$, $p = 0.009$, but higher on WMU, $r = .22$, $p < 0.001$, with no significant differences on other specific factors. The correlational pattern was strongly similar when computed within the structural framework versus when factor scores were exported and correlations were computed externally, except that the relations of fluency with age and sex, while directionally the same as those above and similar in value, were now significant.

Discussion

The goal of this study was to evaluate a framework for EF. A unique contribution of this study is the comprehensiveness with which EFs were measured to encompass different

conceptualizations across different academic literatures, in latent variable fashion, in a large sample in a key school-age range. The best-fitting model included a common EF factor in addition to five distinguishable factors: Working Memory as Span/Manipulate with Planning (WM-S/M and Planning); Working Memory as Updating (WM-U); Generative Fluency; Self-Regulated Learning (SRL); and Metacognition (MCOG).

Results supported the hypothesized framework in that: (a) most indicators had adequate loadings on their hypothesized factor; (b) the bifactor model showed a better fit than a correlated factors model; (c) the SRL and MCOG factors showed higher loadings on their specific factor than that of common EF; (d) relations with demographic variables were generally predictable, particularly with regard to the common EF factor. However, on a number of other points, the originally hypothesized framework was not supported: (a) the indicators chosen did not allow for the separate measurement of MCOG and Behavioral Self-Regulation; (b) correlations among factors in Models 1a, 1b, and 2a were too strongly correlated; (c) indicators of WM-U showed unexpectedly strong *specific* loadings in the bifactor context; (d) latent factors of both Shifting and Inhibition were not identifiable outside of the common EF factor; and (e) relationships of the final specific factors with covariates were more sparse, and weaker, than anticipated.

The structural pattern of this study is partially consistent with a number of other factor analytic studies, though the number and range of EF-related constructs considered here is larger than prior work. We were unable to replicate either the preschool structure of EF as a unitary factor, nor the well-known triumvirate of Shifting, Updating, and Inhibition (Miyake et al., 2000). Prior attempts to reproduce this three-factor structure, even when expressly attempted, often show some inconsistency. For example, in van der Sluis, de Jong, and van der Leij (2007), an inhibition factor was not identified; in van der Ven, Kriesbergen, Boom, and Leseman (2012) with younger children, an inhibition and shifting factor were combined. In twin studies, Friedman et al. (2008) found that inhibition correlated robustly ($r = .73$) with both shifting and updating; in their hierarchical genetic analysis, inhibition had a perfect correlation with Common EF, and in nested (bifactor) analyses, their indicators of inhibition were subsumed in Common EF. Lee et al. (2013) also found that factors of shifting and inhibition were generally combined from ages 6 to 15 (when the three factors were separable). Finally, where studies have clearly considered working memory from the perspective of both updating (e.g., n-back) and storage/maintenance (e.g., listening recall) (Engelhardt et al., 2015), these are found to be separable, at least in children. This issue has been previously considered in meta-analyses that show quite weak correspondence between these two types of measures (Redick & Lindsey, 2013) whereas latent variable models show more overlap (Schmiedek et al., 2009; 2014); although unlike Engelhardt et al. (2015), these studies were with adults. The present results are intriguing but there is a strong need for more work regarding the separation of these factors theoretically and empirically, as well as their relation to outcomes.

How Unitary is EF?

The present study is unable to conclude that indicators of EF are interchangeable and each contributes similarly to a unitary factor; it is also unable to conclude that hypothesized

factors are separable and non-overlapping. That neither is clearly true supports the "separate but related" perspective on EF (Miyake et al., 2000). However, the pattern of structural relations in this study helps to clarify which posited EFs are more "separable" and which are more "related".

Separable EFs—There was a clear distinction among latent variables of: (a) SRL; (b) MCOG; and of (c) several performance measures of EF. That these types of measures arise from more or less different literatures is interesting. Because evaluation of the relationships among these different kinds of variables is somewhat confounded by measurement, it is unclear what drives their differentiation – their method of ascertainment or a fundamental difference in constructs. Where all indicators share a common platform (e.g., all computerized, or all timed, or all experimental, or all from the same co-normed battery), separation may be easier, but combining methods is very common in clinical settings. In single-group predictive or causal-comparative studies, different types of measures contribute differentially and in complementary fashion (e.g., Richardson, Abraham, & Bond, 2012; Gerst, Cirino, Fletcher, & Yoshida, 2017). This issue of methodological variance may have played a role in other studies as well. For example, in Lee et al. (2013), one of their factors (working memory) consisted of accuracy scores, whereas the other factor (combining switch/inhibit) was predominantly assessed with response-time measures. However, a critical finding in the present study was that a contrastive model that *only* considered mode of administration was a substantially worse fit to the data than the hypothesized model (see Table 2), suggesting that the potential viability of the specific EF constructs identified here is not solely due to measurement. More directed structural studies, along with future validity studies, are warranted to sort out the relative utility of this EF disaggregation by measurement type versus construct.

Common EF—However separable EF factors may be, the results of this study suggest there is also some value in considering EF from a more general perspective. A common EF has potential to (literally) unite different and sometimes competing theoretical conceptualizations of EF. To the extent that a common EF can be reliably and robustly measured, it may be possible to choose indicators that most closely tap into it. For example, in the present study, measures of Shifting, Working Memory (as span/manipulation), Planning, and Generative Fluency had robust loadings on the primary EF factor. Even at the specific level, the combination of WM(S/M) with Planning also makes sense conceptually since the act of planning requires simultaneous representation and manipulation of potential action sequences, along with a judgement as to their relative success. Future studies may attempt to ascertain how minimally the common EF factor can be assessed without losing structural identification or predictive power.

The relatively weak interrelationships among the indicator variables (median $r = .15$, see Table S2) were initially unexpected, and prompted a secondary review of prior factor analytic studies. For example, the ten tasks in Wiebe et al. (2008) had a median intercorrelation of $r = .28$; for the nine tasks in Miyake et al. (2000), $r = .13$; for the nine tasks in Lee et al. (2013), $r = .33$ (without partialling their control tasks); and for the 14 tasks in Lehto et al. (2003), $r = .23$. This range is especially noteworthy as those studies span the

age range, utilize different combinations of measures, and come to somewhat different conclusions regarding the structure of EF.

Capitalizing on Knowledge Related to EF

The purpose of EF is to manage goal-directed behavior. Data that speak to the way in which EF can be used for this purpose come from at least three types of studies: (a) structure; (b) relations to other variables; and (c) mechanism and intervention.

Structure—The framework of the present study is fundamentally about how the structure of EF broadly construed can be aggregated versus disaggregated in this age/grade range. Thus structure describes EF; other forms of description include the identification of brain regions implicated in EF performance, or how groups of patients or other subgroups of individuals differ from one another on one or more aspects of EF. Description is key for complex or difficult-to-grasp constructs, which is clearly apt for EF (it is difficult to discuss "the EF, the whole EF, and nothing but the EF"). In the present study we chose to focus the impact on synthesizing diverse ways of referencing this construct. What remains less clear is whether the pattern of loadings found in this study holds up in other populations, particularly in neurodevelopmental disorders or acquired neurological disorders – given that some of these are explicitly associated with or defined by changes in the brain. It is also unclear whether different sets of measures would produce the same set of broad and specific factors. A firmer understanding of the structure of EF is relevant to its practical utility, and can be quite useful in enhancing communication within or across fields, much as the study of Miyake et al. (2000) focused discussion of EF around their three core components of Shifting, Inhibition, and Updating. We view the present study as expanding discussion beyond these three factors in the context of one another.

Relations to Other Variables—In the present study, we focused on demographic characteristics. We found a predictable, positive relationship of the common EF factor with age ($r = .38$) which is robust, even though the age range, albeit important, was not large. This is consistent with prior studies that show relation of EF with age (Engelhardt et al., 2015; Wiebe et al., 2008). The common EF factor also showed predictable (negative) relationships with economic disadvantage (Raver, 2004; Raver et al., 2013), and with limited English proficiency (Abedi, 2002). It is not particularly surprising that girls were rated as having stronger MCOG than boys, given the well-known elevated rates of attentional and related difficulties for boys relative to girls (DuPaul, et al., 1997; Gershon & Gershon, 2002). Although there is little data with which to compare current results, given the need for English proficiency on an (English) verbal fluency task, it is also not particularly surprising that a negative correlation of performance and limited English proficiency was noted ($r = -.26$).

It is less clear why there was a significant negative relation of age and MCOG ($r = -.19$), implying that older students have less developed skills in this area, since the expectation would be that an increase in age produces an increase in meta-cognition. It should be noted that the absolute value of this relationship was not strong, and that the raw indicator scores showed no significant relation with age; also the relation with age was reduced in absolute

size when the sample was not weighted. Notably, the MCOG factor is derived to be uncorrelated with all other latent variables, both specific and common, and therefore, its relationship is closer to a beta weight then a zero-order correlation, since any shared variance is attributed to the common EF factor. A second less expected (directionally opposite) relation was found between Working Memory (as updating) for students with limited English proficiency ($r = .22$), where such students exhibited stronger performance. The raw scores produced somewhat similar patterns. It is possible to interpret this albeit small effect in light of the "bilingual advantage" literature, where it has been suggested that bilingual individuals have improved EF performances (Bialystock, 1999; see Morton & Harper, 2007, and Paap & Greenberg, 2013), but two caveats may be in order. First, the bilingual advantage is typically associated with individuals who excel at both languages, whereas the students in the present study by definition do not; some of the students are bilingual in the sense that they have skills in both languages, but they are still in the process of learning English, even though their instruction occurs in English. Second, all the students with limited English proficiency were in the fourth grade and all were struggling readers.

Mechanism and Intervention—Beyond structure/description and relation to other variables, it is critical to understand how and/or why EF operates to manage goal-directed behavior. Descriptive studies, even those with large sample sizes and theoretically driven and comprehensive measures, cannot provide information relevant to mechanism and causal direction. In contrast, experimental studies, including intervention studies, allows for a recognition of key ingredients and how they might be implemented and for whom, particularly within populations where EF is compromised, or where it is more strongly related to outcomes – for example, academic achievement (Best, Miller, & Naglieri, 2011; Biederman et al., 2004). In their review, Jacob and Parkinson (2015) found clear and consistent relationships between EF and academic achievement, but interventions designed to improve EF had little generalizability for improving achievement. Although the structural framework of EF and how it relates to other variables can set the parameters for other types of studies, an understanding of this mechanism is needed for (and is testable by) interventions designed to affect that mechanism (see for example, recent work by Fuchs, Fuchs, & Malone, 2016).

Limitations and Conclusions

One potential limitation is how the nature of the sampling plan impacts generalizability. Students with reading difficulty were oversampled (purposefully). Weights were applied to individuals' data so that the overall contributions from the oversampled students were in proportion to their distribution in the schools as a whole. However, there is no rationale to hypothesize a different *pattern* of relationships based on *level* of performance. In addition, weighting is common in large scale national samples to increase verisimilitude to the population. A second limitation is that while it would have been ideal to administer all measures to all students, this was not practicable. Indeed, it would also be ideal to administer *more* measures (e.g., wider coverage of working memory). Similarly, including both performance and rating measures of metacognition would have been interesting, particularly in light of the weak correlations between these kinds of measures with regard to EF generally (e.g., Gerst et al., 2017; Toplak, West, and Stanovich, 2013). In this study, we

opted to increase construct coverage at the cost of covariance coverage. However, the missingness was planned (by design) via randomization and so did not bias structural loadings (Rhemtulla & Little, 2012). Finally, EF structure was not evaluated in a multilevel context; however, accounting for nesting effects would be important to address where EF is used to predict outcomes, particularly those that are instructed (e.g., reading).

In spite of these limitations, the present study benefitted from a large, diverse sample, and the statistical approach was inclusive and confirmatory, examining both correlated factors and bifactor models. The best-fitting model was a bifactor model with a common EF factor, and specific factors of (a) Working Memory (as storage/maintenance) combined with Planning; (b) Working Memory (as updating); (c) Generative Fluency; (d) Self-Regulated Learning; and (e) Metacognition. This model speaks to both the robustness of a general factor of EF, but also to the potential separable impact of specific EF factors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Abedi J. Standardized achievement tests and English language learners: Psychometrics issues. Educational Assessment. 2002; 8(3):231–257. DOI: 10.1207/S15326977EA0802_02

Allan NP, Lonigan CJ. Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. Developmental psychology. 2011; 47(4):905. http://dx.doi.org/10.1037/a0023748. [PubMed: 21553957]

Alvarez JA, Emory E. Executive function and the frontal lobes: a meta-analytic review. Neuropsychology Review. 2006; 16(1):17–42. DOI: 10.1007/s11065-006-9002-x [PubMed: 16794878]

Anderson VA, Anderson P, Northam E, Jacobs R, Catroppa C. Development of executive functions through late childhood and adolescence in an Australian sample. Developmental Neuropsychology. 2001; 20(1):385–406. DOI: 10.1207/S15326942DN2001_5 [PubMed: 11827095]

Author. The Contextual Learning Scale. 2012

Baddeley AD, Allen RJ, Hitch GJ. Binding in visual working memory: The role of the episodic buffer. Neuropsychologia. 2011; 49(6):1393–1400. DOI: 10.1016/j.neuropsychologia.2010.12.042 [PubMed: 21256143]

Barkley RA. Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. Psychological Bulletin. 1997; 121(1):65–94. DOI: 10.1037/0033-2909.121.1.65 [PubMed: 9000892]

Barkley, RA., editor. Attention-deficit hyperactivity disorder: A Handbook for Diagnosis and Treatment. New York, NY: Guilford Publications; 2014.

Berch DB, Krikorian R, Huha EM. The Corsi block-tapping task: Methodological and theoretical considerations. Brain and Cognition. 1998; 38(3):317–338. DOI: 10.1006/brcg.1998.1039 [PubMed: 9841789]

Best JR, Miller PH, Jones LL. Executive functions after age 5: Changes and correlates. Developmental Review. 2009; 29(3):180–200. DOI: 10.1016/j.dr.2009.05.002 [PubMed: 20161467]

Best JR, Miller PH, Naglieri JA. Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. Learning and Individual Differences. 2011; 21(4):327–336. DOI: 10.1016/j.lindif.2011.01.007 [PubMed: 21845021]

Bialystok E. Cognitive complexity and attentional control in the bilingual mind. Child Development. 1999; 70(3):636–644. DOI: 10.1111/1467-8624.00046

Biederman J, Monuteaux MC, Doyle AE, Seidman LJ, Wilens TE, Ferrero F, Faraone SV. Impact of executive function deficits and attention-deficit/hyperactivity disorder (ADHD) on academic outcomes in children. Journal of Consulting and Clinical Psychology. 2004; 72(5):757. http://dx.doi.org/10.1037/0022-006X.72.5.757. [PubMed: 15482034]

Blair C. School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. American Psychologist. 2002; 57(2):111. http://dx.doi.org/10.1037/0003-066X.57.2.111. [PubMed: 11899554]

Borkowski, JG., Chan, LK., Muthukrishna, N. 1. A Process-Oriented Model of Metacognition: Links Between Motivation and Executive Functioning. In: Schraw, G., Impara, JC., editors. Issues in the Measurement of Metacognition. Lincoln, NE: Buros Institute of Mental Measurements; 2000.

Bronson, MB., Goodson, BD., Layzer, JI., Love, JM. Child Behavior Rating Scale. Cambridge: MA: Abt Associates; 1990.

Brydges CR, Fox AM, Reid CL, Anderson M. The differentiation of executive functions in middle and late childhood: A longitudinal latent-variable analysis. Intelligence. 2014; 47:34–43. DOI: 10.1016/j.intell.2014.08.010

Burnham KP, Anderson DR. Multimodel inference understanding AIC and BIC in model selection. Sociological Methods & Research. 2004; 33(2):261–304. DOI: 10.1177/0049124104268644

Canivez GL. Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. School Psychology Quarterly. 2014; 29(1):38.doi: 10.1037/spq0000032 [PubMed: 23895320]

Cleary TJ. The development and validation of the self-regulation strategy inventory—self-report. Journal of School Psychology. 2006; 44(4):307–322. DOI: 10.1016/j.jsp.2006.05.002

Cowan N. What are the differences between long-term, short-term, and working memory? Progress in brain research. 2008; 169:323–338. https://doi.org/10.1016/S0079-6123(07)00020-9. [PubMed: 18394484]

Delis, D., Kaplan, E., Kramer, J. Delis-Kaplan Executive Function Scale. San Antonio, TX: The Psychological Corporation; 2001.

Duckworth AL, Kern ML. A meta-analysis of the convergent validity of self-control measures. Journal of Research in Personality. 2011; 45(3):259–268. DOI: 10.1016/j.jrp.2011.02.004 [PubMed: 21643479]

DuPaul GJ, Power TJ, Anastopoulos AD, Reid R, McGoey KE, Ikeda MJ. Teacher ratings of attention deficit hyperactivity disorder symptoms: Factor structure and normative data. Psychological Assessment. 1997; 9(4):436.doi: 10.1037/pas0000166

Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Structural Equation Modeling. 2001; 8(3):430–457. DOI: 10.1207/S15328007SEM0803_5

Engle RW. Working memory capacity as executive attention. Current Directions in Psychological Science. 2002; 11(1):19–23. DOI: 10.1111/1467-8721.00160

Engelhardt LE, Briley DA, Mann FD, Harden KP, Tucker-Drob EM. Genes unite executive functions in childhood. Psychological Science. 2015; 26(8):1151–1163. DOI: 10.1177/0956797615577209 [PubMed: 26246520]

Floyd RG, Bergeron R, Hamilton G, Parra GR. How do executive functions fit with the Cattell–Horn–Carroll model? Some evidence from a joint factor analysis of the Delis–Kaplan executive function system and the Woodcock–Johnson III tests of cognitive abilities. Psychology in the Schools. 2010; 47(7):721–738. DOI: 10.1002/pits.20500

Friedman NP, Miyake A, Robinson JL, Hewitt JK. Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: a behavioral genetic analysis. Developmental Psychology. 2011; 47(5):1410.doi: 10.1037/a0023750 [PubMed: 21668099]

Friedman NP, Miyake A, Young SE, DeFries JC, Corley RP, Hewitt JK. Individual differences in executive functions are almost entirely genetic in origin. Journal of Experimental Psychology: General. 2008; 137(2):201.doi: 10.1037/0096-3445.137.2.201 [PubMed: 18473654]

Fuchs LS, Fuchs D, Malone AS. Multilevel Response-to-Intervention Prevention Systems: Mathematics Intervention at Tier 2. Handbook of Response to Intervention. 2016; :309–328. DOI: 10.1007/978-1-4899-7568-3_18

Garon N, Bryson SE, Smith IM. Executive function in preschoolers: a review using an integrative framework. Psychological bulletin. 2008; 134(1):31.doi: 10.1037/0033-2909.134.1.31 [PubMed: 18193994]

Gershon J, Gershon J. A meta-analytic review of gender differences in ADHD. Journal of attention disorders. 2002; 5(3):143–154. DOI: 10.1177/108705470200500302 [PubMed: 11911007]

Gerst EH, Cirino PT, Fletcher JM, Yoshida H. Cognitive and behavioral rating measures of executive function as predictors of academic outcomes in children. Child Neuropsychology. 2017; 23(4):381–407. DOI: 10.1080/09297049.2015.1120860 [PubMed: 26678705]

Gioia, GA., Isquith, PA., Guy, SC., Kenworthy, L. Behavior Rating Inventory of Executive Function. Odessa, FL: Psychological Assessment Resources, Inc; 2000.

Gioia GA, Isquith PK, Retzlaff PD, Espy KA. Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample. Child Neuropsychology. 2002; 8(4):249–257. DOI: 10.1076/chin.8.4.249.13513 [PubMed: 12759822]

Hackman DA, Farah MJ. Socioeconomic status and the developing brain. Trends in Cognitive Sciences. 2009; 13(2):65–73. DOI: 10.1016/j.tics.2008.11.003 [PubMed: 19135405]

Hackman DA, Gallop R, Evans GW, Farah MJ. Socioeconomic status and executive function: developmental trajectories and mediation. Developmental Science. 2015; 18(5):686–702. DOI: 10.1111/desc.12246 [PubMed: 25659838]

Hofmann W, Schmeichel BJ, Baddeley AD. Executive functions and self-regulation. Trends in Cognitive Sciences. 2012; 16(3):174–180. DOI: 10.1016/j.tics.2012.01.006 [PubMed: 22336729]

Holland, PW., Thayer, DT. Notes on the use of log-linear models for fitting discrete probability distributions (Technical Report 87–79). Princeton, NJ: Educational Testing Service; 1987.

Holzinger KJ, Swineford F. The bifactor method. Psychometrika. 1937; 2(1):41–54. DOI: 10.1007/BF02287965

Hu L, Bentler M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6(1):1–55. http://dx.doi.org/10.1080/10705519909540118.

Huizinga M, Dolan CV, van der Molen MW. Age-related change in executive function: Developmental trends and a latent variable analysis. Neuropsychologia. 2006; 44(11):2017–2036. DOI: 10.1016/j.neuropsychologia.2006.01.010 [PubMed: 16527316]

Hull R, Martin RC, Beier ME, Lane D, Hamilton AC. Executive function in older adults: a structural equation modeling approach. Neuropsychology. 2008; 22(4):508.doi: 10.1037/0894-4105.22.4.508 [PubMed: 18590362]

Ilkowska M, Engle RW. Working Memory Capacity and Self-Regulation. Handbook of Personality and Self-Regulation. 2010; :263–290. DOI: 10.1002/9781444318111.ch12

Inquisit 3. Seattle, WA: Millisecond Software; 2003.

Jacob R, Parkinson J. The Potential for School-Based Interventions That Target Executive Function to Improve Academic Achievement A Review. Review of Educational Research. 2015; 85(4):512–552. DOI: 10.3102/0034654314561338

Jonides J, Smith EE, Koeppe RA, Awh E, Minoshima S, Mintun MA. Spatial working-memory in humans as revealed by PET. Nature. 1993; 363:623–625. http://dx.doi.org/10.1038/363623a0. [PubMed: 8510752]

Jurado MB, Rosselli M. The Elusive Nature of Executive Functions: A Review of our Current Understanding. Neuropsychology Review. 2007; 17(3):213–233. DOI: 10.1007/s11065-007-9040-z [PubMed: 17786559]

Kaller CP, Rahm B, Köstering L, Unterrainer JM. Reviewing the impact of problem structure on planning: a software tool for analyzing tower tasks. Behavioural Brain Research. 2011; 216(1):1–8. DOI: 10.1016/j.bbr.2010.07.029 [PubMed: 20723568]

Kaller CP, Unterrainer JM, Stahl C. Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. Psychological Assessment. 2012; 24(1):46. http://dx.doi.org/10.1037/a0025174. [PubMed: 21859218]

Kane MJ, Hambrick DZ, Tuholski SW, Wilhelm O, Payne TW, Engle RW. The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. Journal of Experimental Psychology: General. 2004; 133(2):189. http://dx.doi.org/10.1037/0096-3445.133.2.189. [PubMed: 15149250]

Kirchner WK. Age differences in short-term retention of rapidly changing information. Journal of Experimental Psychology. 1958; 55(4):352. http://dx.doi.org/10.1037/h0043688. [PubMed: 13539317]

Kline, RB. Principles and practice of structural equation modeling 2011. New York: Guilford Press; 2010.

Lee K, Bull R, Ho RM. Developmental changes in executive functioning. Child Development. 2013; 84(6):1933–1953. DOI: 10.1111/cdev.12096 [PubMed: 23550969]

Lehto JE, Juujärvi P, Kooistra L, Pulkkinen L. Dimensions of executive functioning: Evidence from children. British Journal of Developmental Psychology. 2003; 21(1):59–80. DOI: 10.1348/026151003321164627

Levine B, Robertson IH, Clare L, Carter G, Hong J, Wilson BA, Stuss DT. Rehabilitation of executive functioning: An experimental–clinical validation of goal management training. Journal of the International Neuropsychological Society. 2000; 6(03):299–312. [PubMed: 10824502]

Levine B, Schweizer TA, O'Connor C, Turner G, Gillingham S, Stuss DT, Manly T, Robertson IH. Rehabilitation of executive functioning in patients with frontal lobe brain damage with goal management training. Frontiers in Human Neuroscience. 2011; 5(9):1–9. DOI: 10.3389/fnhum.2011.00009 [PubMed: 21283556]

Lezak, MD. Neuropsychological Assessment. 3rd. New York, NY: Oxford University Press; 1995.

Lijffijt M, Kenemans JL, Verbaten MN, van Engeland H. A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: deficient inhibitory motor control? Journal of Abnormal Psychology. 2005; 114(2):216–222. http://dx.doi.org/10.1037/0021-843X.114.2.216. [PubMed: 15869352]

Little RJ. A test of missing completely at random for multivariate data with missing values. Journal of the American Statistical Association. 1988; 83(404):1198–1202. DOI: 10.1080/01621459.1988.10478722

Little TD, Jorgensen TD, Lang KM, Moore EWG. On the joys of missing data. Journal of pediatric psychology. 2014; 39(2):151–162. https://doi.org/10.1093/jpepsy/jst048. [PubMed: 23836191]

Logan GD, Cowan WB. On the ability to inhibit thought and action: A theory of an act of control. Psychological Review. 1984; 91(3):295–327. http://dx.doi.org/10.1037/0033-295X.91.3.295.

Luciana M, Nelson CA. Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: performance in 4-to 12-year-old children. Developmental Neuropsychology. 2002; 22(3):595–624. DOI: 10.1207/S15326942DN2203_3 [PubMed: 12661972]

MacGinitie, W., MacGinitie, R., Maria, K., Dreyer, LG., Hughes, KE. Gates-MacGinitie Reading Tests (GMRT). Fourth. Itasca, IL: Riverside Publishing; 2000.

McGrew, KS., Woodcock, RW. Woodcock-Johnson III: Normative Update Technical Manual. Riverside Publications; 2001.

Midgley, C., Maehr, ML., Hicks, L., Roeser, R., Urdan, T., Anderman, E., Middleton, M. Patterns of Adaptive Learning Survey (PALS). Ann Arbor, MI: Center for Leadership and Learning; 1996.

Milner B, Petrides M. Behavioural effects of frontal-lobe lesions in man. Trends in Neurosciences. 1984; 7(11):403–407. DOI: 10.1016/S0166-2236(84)80143-5

Miyake A, Friedman NP. The nature and organization of individual differences in executive functions four general conclusions. Current Directions in Psychological Science. 2012; 21(1):8–14. DOI: 10.1177/0963721411429458 [PubMed: 22773897]

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cognitive Psychology. 2000; 41(1):49–100. DOI: 10.1006/cogp.1999.0734 [PubMed: 10945922]

Morton JB, Harper SN. What did Simon say? Revisiting the bilingual advantage. Developmental Science. 2007; 10(6):719–726. DOI: 10.1111/j.1467-7687.2007.00623.x [PubMed: 17973787]

Moses, T., von Davier, AA. Stinson, W., Westerlund, E., editors. Using PROC GENMOD for loglinear smoothing. NorthEast SAS Users Group, Inc seventeenth annual conference proceedings. 2004. Retrieved January 12, 2016, from http://www.lexjansen.com/nesug/nesug04/an/an07.pdf

Muthén B, Kaplan D, Hollis M. On structural equation modeling with data that are not missing completely at random. Psychometrika. 1987; 52(3):431–462. DOI: 10.1007/BF02294365

Muthén, LK., Muthén, BO. Mplus (Version 7.0). Los Angeles, CA: 2012.

National Center for Education Statistics. English learners in public schools. 2016. Accessed electronically July 1, 2016 at http://nces.ed.gov/programs/coe/pdf/coe_cgf.pdf

Oosterlaan J, Logan GD, Sergeant JA. Response inhibition in AD/HD, CD, comorbid AD/HD+ CD, anxious, and control children: a meta-analysis of studies with the stop task. Journal of Child Psychology and Psychiatry. 1998; 39(03):411–425. [PubMed: 9670096]

Paap KR, Greenberg ZI. There is no coherent evidence for a bilingual advantage in executive processing. Cognitive Psychology. 2013; 66(2):232–258. DOI: 10.1016/j.cogpsych.2012.12.002 [PubMed: 23370226]

Pagulayan K, Busch RM, Medina KL, Bartok JA, Krikorian R. Developmental normative data for the Corsi block-tapping task. Journal of Clinical and Experimental Neuropsychology. 2006; 28(6): 1043–1052. DOI: 10.1080/13803390500350977 [PubMed: 16822742]

Pickering, SJ., Gathercole, SE. Working Memory Test Battery for Children. London: Psychological Corporation Europe; 2001.

Pintrich, P. The role of goal orientation in self-regulated learning. In: Boekaerts, M.Pintrich, PR., Zeidner, M., editors. Handbook of Self-Regulation. San Diego, CA, US: Academic Press; 2000. p. 451-502.

Pintrich P. A conceptual framework for assessing motivation and self-regulated learning in college students. Educational Psychology Review. 2004; 16(4):385–407. DOI: 10.1007/s10648-004-0006-x

Pintrich PR, De Groot EV. Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology. 1990; 82(1):33. http://dx.doi.org/10.1037/0022-0663.82.1.33.

Posner MI, Rothbart MK. Developing mechanisms of self-regulation. Development and Psychopathology. 2000; 12(03):427–441. [PubMed: 11014746]

Raver CC. Placing emotional self-regulation in sociocultural and socioeconomic contexts. Child Development. 2004; 75(2):346–353. DOI: 10.1111/j.1467-8624.2004.00676.x [PubMed: 15056189]

Raver CC, Blair C, Willoughby M, The Family Life Project Key Investigators. Poverty as a predictor of 4-year-olds' executive function: New perspectives on models of differential susceptibility. Developmental Psychology. 2013; 49(2):292.doi: 10.1037/a0028343 [PubMed: 22563675]

Raykov T, Hancock GR. Examining change in maximal reliability for multiple‐ component measuring instruments. British Journal of Mathematical and Statistical Psychology. 2005; 58(1):65–82. DOI: 10.1348/000711005X38753 [PubMed: 15969840]

Redick TS, Calvo A, Gay CE, Engle RW. Working memory capacity and go/no-go task performance: Selective effects of updating, maintenance, and inhibition. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2011; 37(2):308. http://dx.doi.org/10.1037/a0022216.

Redick TS, Lindsey DR. Complex span and n-back measures of working memory: A meta-analysis. Psychonomic Bulletin & Review. 2013; 20(6):1102–1113. DOI: 10.3758/s13423-013-0453-9 [PubMed: 23733330]

Reise SP. The rediscovery of bifactor measurement models. Multivariate Behavioral Research. 2012; 47(5):667–696. DOI: 10.1080/00273171.2012.715555 [PubMed: 24049214]

Rhemtulla M, Hancock GR. Planned missing data designs in educational psychology research. Educational Psychologist. 2016; 51(3–4):305–316. http://dx.doi.org/10.1080/00461520.2016.1208094.

Rhemtulla M, Little TD. Planned missing data designs for research in cognitive development. Journal of Cognition and Development. 2012; 13(4):425–438. http://dx.doi.org/10.1080/15248372.2012.717340.

Richardson M, Abraham C, Bond R. Psychological correlates of university students' academic performance: a systematic review and meta-analysis. Psychological Bulletin. 2012; 138(2): 353.doi: 10.1037/a0026838 [PubMed: 22352812]

Roberts RJ Jr, Pennington BF. An interactive framework for examining prefrontal cognitive processes. Developmental Neuropsychology. 1996; 12(1):105–126. DOI: 10.1080/87565649609540642

Rubin DB. Inference and missing data. Biometrika. 1976; :581–592. DOI: 10.2307/2335739

Schachar R, Mota VL, Logan GD, Tannock R, Klim P. Confirmation of an inhibitory control deficit in attention-deficit/hyperactivity disorder. Journal of Abnormal Child Psychology. 2000; 28(3):227–235. DOI: 10.1023/A:1005140103162 [PubMed: 10885681]

Schmiedek F, Hildebrandt A, Lövdén M, Wilhelm O, Lindenberger U. Complex span versus updating tasks of working memory: The gap is not that deep. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2009; 35(4):1089.doi: 10.1037/a0015730

Schmiedek F, Lövdén M, Lindenberger U. Younger adults show long-term effects of cognitive training on broad cognitive abilities over 2 years. Developmental Psychology. 2014; 50(9):2304.doi: 10.1037/a0037388 [PubMed: 25019946]

Shallice T. Specific Impairments of Planning. Philosophical Transactions of the Royal Society of London. B, Biological Sciences. 1982; 298(1089):199–209. DOI: 10.1098/rstb.1982.0082 [PubMed: 6125971]

Solanto MV, Abikoff H, Sonuga-Barke E, Schachar R, Logan GD, Wigal T, Turkel E. The ecological validity of delay aversion and response inhibition as measures of impulsivity in AD/HD: a supplement to the NIMH multimodal treatment study of AD/HD. Journal of Abnormal Child Psychology. 2001; 29(3):215–228. DOI: 10.1023/A:1010329714819 [PubMed: 11411784]

Speece DL, Ritchey KD, Silverman R, Schatschneider C, Walker CY, Andrusik KN. Identifying children in middle childhood who are at risk for reading problems. School Psychology Review. 2010; 39(2):258. [PubMed: 21472039]

Spreen, O., Strauss, E. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. New York: Oxford University Press; 1991.

Stroop J. Studies of inference in serial verbal reactions. Journal of Experimental Psychology. 1935; 18:643–662. http://dx.doi.org/10.1037/h0054651.

Stroud, KC., Reynolds, CR. The School Motivation and Learning Strategies Inventory. Los Angeles: Western Psychological Services; 2006.

Stuss DT. Functions of the frontal lobes: Relation to executive functions. Journal of the International Neuropsychological Society. 2011; 17(5):759–765. DOI: 10.1017/S1355617711000695 [PubMed: 21729406]

Swanson JM, Schuck S, Porter MM, Carlson C, Hartman CA, Sergeant JA, Wigal T. Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. The International journal of educational and psychological assessment. 2012; 10(1):51. [PubMed: 26504617]

Tranel, D., Anderson, SW., Benton, A. Development of the concept of "executive function" and its relationship to the frontal lobes. In: Boller, F., Grafman, J., editors. Handbook of Neuropsychology. Amsterdam, The Netherlands: Elsevier Science BV; 1994. p. 125-148.

Toplak ME, West RF, Stanovich KE. Do performance-based measures and ratings of executive function assess the same construct? Journal of Child Psychology and Psychiatry. 2013; 54(2): 131–143. DOI: 10.1111/jcpp.12001 [PubMed: 23057693]

van der Sluis S, de Jong PF, van der Leij A. Executive functioning in children, and its relations with reasoning, reading, and arithmetic. Intelligence. 2007; 35(5):427–449. DOI: 10.1016/j.intell.2006.09.001

van der Ven SHG, Kroesbergen EH, Boom J, Leseman PPM. The development of executive functions and early mathematics: A dynamic relationship. British Journal of Educational Psychology. 2012; 82(1):100–119. DOI: 10.1111/j.2044-8279.2011.02035.x [PubMed: 22429060]

Welsh MC, Pennington BF, Groisser DB. A normative-developmental study of executive function: A window on prefrontal function in children. Developmental Neuropsychology. 1991; 7(2):131–149. DOI: 10.1080/87565649109540483

Welsh M, Peterson E. Issues in the conceptualization and assessment of hot executive functions in childhood. Journal of the International Neuropsychological Society. 2014; 20(02):152–156. DOI: 10.1017/S1355617713001379 [PubMed: 24468077]

Wiebe SA, Espy KA, Charak D. Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. Developmental Psychology. 2008; 44(2):575.doi: 10.1037/0012-1649.44.2.575 [PubMed: 18331145]

Wigfield A, Guthrie JT. Relations of children's motivation for reading to the amount and breadth or their reading. Journal of Educational Psychology. 1997; 89(3):420. http://dx.doi.org/10.1037/0022-0663.89.3.420.

Willcutt EG, Pennington BF, Olson RK, Chhabildas N, Hulslander J. Neuropsychological analyses of comorbidity between reading disability and attention-deficit hyperactivity disorder: In search of the common deficit. Developmental Neuropsychology. 2005; 27(1):35–78. http://dx.doi.org/10.1207/s15326942dn2701_3. [PubMed: 15737942]

Willoughby MT, Blair CB, Wirth RJ, Greenberg M. The measurement of executive function at age 3 years: psychometric properties and criterion validity of a new battery of tasks. Psychological Assessment. 2010; 22(2):306.doi: 10.1037/a0018708 [PubMed: 20528058]

Willoughby MT, Pek J, Blair CB. Measuring executive function in early childhood: A focus on maximal reliability and the derivation of short forms. Psychological Assessment. 2013; 25(2):664.doi: 10.1037/a0031747 [PubMed: 23397928]

Woodcock, RW., McGrew, KS., Mather, N. The Woodcock-Johnson III Tests of Achievement. Illinois: Riverside Publishing; 2001.

Yakovlev, PI., Lecours, AR. The myelogenetic cycles of regional maturation of the brain. In: Minkowski, A., editor. Regional Development of the Brain in Early Life. Oxford: Blackwell; 1967. p. 3-70.

Ylvisaker M, Feeney T. Executive functions, self-regulation, and learned optimism in paediatric rehabilitation: a review and implications for intervention. Developmental Neurorehabilitation. 2002; 5(2):51–70. DOI: 10.1080/136384902100004189

Zelazo PD, Carlson SM. Hot and cool executive function in childhood and adolescence: Development and plasticity. Child Development Perspectives. 2012; 6(4):354–360. DOI: 10.1111/j.1750-8606.2012.00246.x

Zelazo PD, Carter A, Reznick JS, Frye D. Early development of executive function: A problem-solving framework. Review of General Psychology. 1997; 1(2):198. http://dx.doi.org/10.1037/1089-2680.1.2.198.

Zimmerman, BJ. Self-Regulated Learning and Academic Achievement. New York, NY: Springer; 1989. Models of self-regulated learning and academic achievement; p. 1-25.

Zimmerman BJ. Self-efficacy: An essential motive to learn. Contemporary Educational Psychology. 2000; 25(1):82–91. DOI: 10.1006/ceps.1999.1016 [PubMed: 10620383]

Zimmerman, BJ., Schunk, DH. Self-regulated learning: An introduction and an overview. In: Zimmerman, BJ., Schunk, DH., editors. Handbook of Self-Regulated Learning and Performance. New York, NY: Routledge; 2011.

Public Significance

Terms such as executive function (EF) are used in many different ways by different scientists as well as different professional and commercial interests. The present study brings together many of the ways that EF is operationalized across these literatures and other arenas. There is more commonality, but also some separability, across these different operationalizations, which has implications for its conceptualization and use.
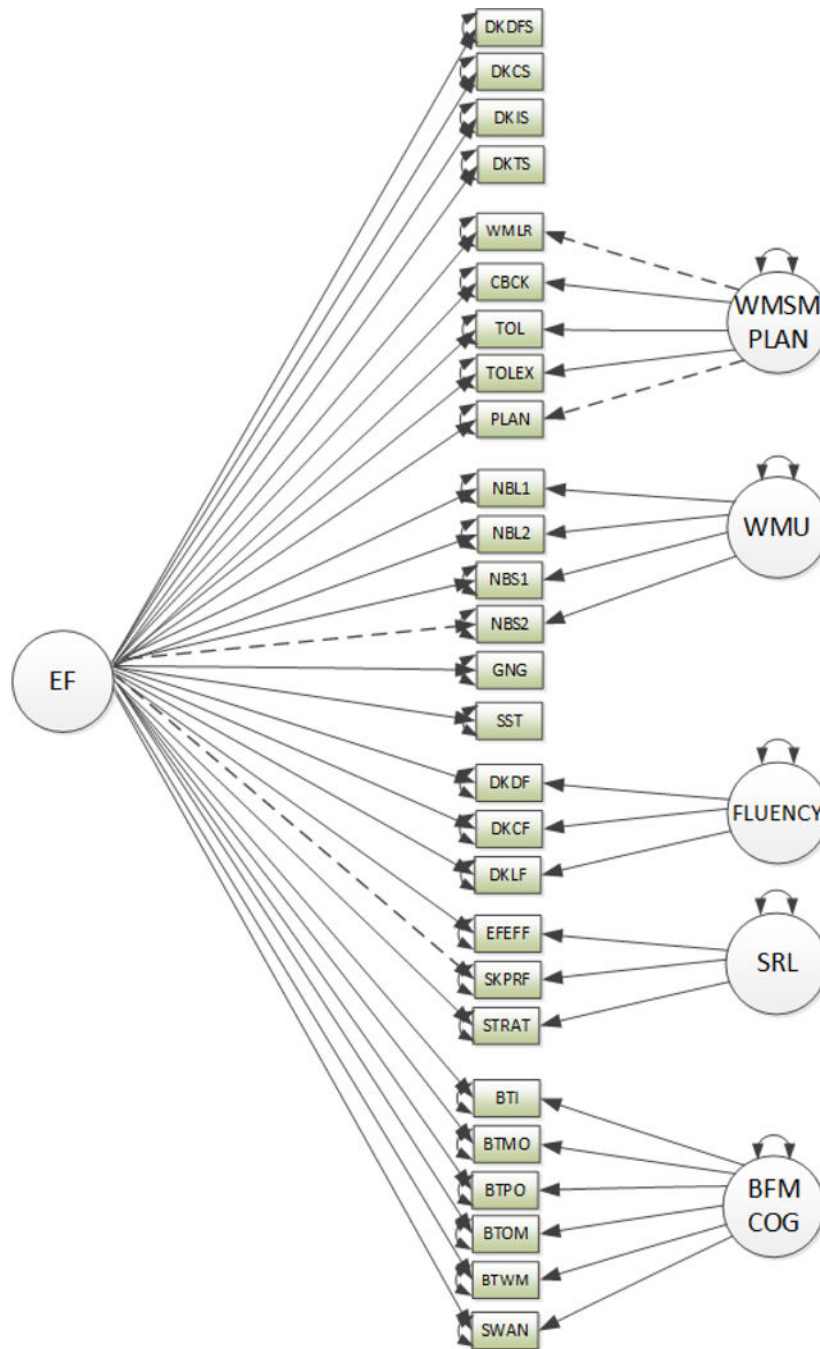
Figure 1. Bifactor model with 5 specific and one general factors (model 4)
Note. DKDFS = D-KEFS Design Fluency Switching; DKCS = D-KEFS Verbal Fluency
Category Switching; DKIS = D-KEFS Color Word Interference Inhibition Switching; DKTS
= D-KEFS Trailmaking Test Letter-Number Switching; WMLR = WMTB-C Listening
Recall; CBCK = Corsi Blocks Backwards Recall; TOL = Tower of London Accuracy;
TOLEX = Tower of London Extra Moves; PLAN = WJ-III Planning; NBL1 = n-back 1
Letter; NBL2 = n-back 2 Letter; NBS1 = n-back 1Shape; NBS2 = n-back 2 Shape; GNG =
Go No-Go; SST = Stop Signal Task; DKDF = D-KEFS Design Fluency Filled+Empty Dots;

DKCF = D-KEFS Verbal Fluency Categories; DKLF = D-KEFS Verbal Fluency Letters; EFEFF = CLS Effort/Self-Efficacy; SKPRF = CLS Skill/Preference; STRAT = CLS Learning Strategies; BTI = BRIEF Initiate; BTMO = BRIEF Monitor; BTPO = BRIEF Planning/Organization; BTOM = BRIEF Organization of Materials; BTWM = BRIEF Working Memory; SWAN = SWAN Inattention.

WM-S/M = Working Memory as Span/Manipulate; WMU= Working Memory as Updating; SRL = Self-Regulated Learning; MCOG = Metacognition.

Solid lines = statistically significant at the $p < 0.05$ level; dashed lines = not statistically significant; rectangles = measured variables; circles = latent variables; double-headed arrows = variances; single-headed arrows = loadings

Table 1

Descriptive statistics and data patterns.

Variable	N	%N	Mean	SD	Skew	Kurt	Rel	Data Patterns 1	2	3	4	5	6
Shifting							0.80						
D-KEFS[a]. Design Switch (max = 9)	420	50	3.98	1.79	0.32	0.16			×		×		×
D-KEFS[a]. Verbal Switch (max = 14)	419	50	5.02	2.95	0.09	-0.43		×				×	
D-KEFS[b]. Inhibit Switch (max = 177)	408	48	91.55	24.97	1.31	1.99				×		×	×
D-KEFS[b]. Trails Switch (max = 240)	410	48	159.26	58.96	0.05	-1.35		×	×	×			
WM-S/M and Plan							0.79						
Working Memory[a] (max = 22)	434	51	11.37	3.70	-0.09	-0.02		×	×	×			
Corsi Blocks[a] (max = 47)	422	50	19.29	9.45	0.10	-0.13				×		×	×
Tower of London[a] (max = 78)	832	98	51.04	15.11	-1.17	1.61		×	×	×	×	×	×
Tower of London[a] (exmov) (max = 4.5)	832	98	1.71	1.05	0.04	-0.78		×	×	×	×	×	×
WJ-III Planning[c] (max = 504)	807	95	500.05	1.76	-0.98	2.66		×	×	×	×	×	×
WMU							0.82						
N-Back[d]. 1 Letter (max = 4.10)	210	25	1.68	1.14	0.26	-0.19			×		×		×
N-Back[d]. 2 Letters (max = 4.10)	210	25	0.80	1.06	1.01	2.01			×		×		×
N-Back[d]. 1 Shape (max = 4.10)	408	48	0.98	0.97	0.49	0.71		×				×	
N-Back[d]. 2 Shapes (max = 2.52)	408	48	0.27	0.68	0.50	0.95		×				×	
Inhibition							0.54						
Cued Go No-Go[d] (max = 4.30)	556	66	3.29	0.65	-0.69	0.49		×	×		×	×	
Stop Signal Test[b] (max = 773)	524	62	319.82	109.24	0.68	0.61		×	×	×	×		×
Fluency							0.77						
D-KEFS[a]. Filled/Empty Dots (max = 26)	827	98	11.22	4.17	0.52	-0.07		×	×	×	×	×	×
D-KEFS[a]. Category Fluency (max = 48)	832	98	24.52	6.78	0.37	0.15		×	×	×	×	×	×
D-KEFS[a]. Letter Fluency (max = 47)	832	98	18.22	7.13	0.48	0.27		×	×	×	×	×	×
Self-Regulation Learning							0.88						

								Data Patterns					
Variable	N	%N	Mean	SD	Skew	Kurt	Rel	1	2	3	4	5	6
Effort/Self-Efficacy[a] (max = 42)	826	98	30.86	7.85	-0.99	0.86		x	x	x	x	x	x
Skill/Preference[a] (max = 15)	826	98	8.33	3.67	-0.25	-0.62		x	x	x	x	x	x
Learning Strategies[a] (max = 54)	827	98	30.88	11.30	-0.06	-0.66		x	x	x	x	x	x
Metacognition							**0.97**						
BRIEF[a]: Initiate (max = 21)	534	63	12.49	4.33	0.33	-0.99		x	x	x	x	x	x
BRIEF[a]: Monitor (max = 30)	534	63	17.46	5.98	0.50	-0.81		x	x	x	x	x	x
BRIEF[a]: Planning/Organizing (max = 30)	534	63	17.46	5.81	0.41	-0.89		x	x	x	x	x	x
BRIEF[a]: Organization (max = 21)	533	63	11.16	4.43	0.80	-0.68		x	x	x	x	x	x
BRIEF[a]: Working Memory (max = 30)	535	63	16.92	6.09	0.55	-0.82		x	x	x	x	x	x
SWAN[a]: Inattention (max = 27)	750	89	0.83	12.70	-0.21	-0.38		x	x	x	x	x	x

Note. D-KEFS = Delis-Kaplan Executive Function System; BRIEF = Behavior Rating Inventory of Executive Function; WJ-III = Woodcock Johnson III; SWAN = Strengths and Weaknesses of Attention-Deficit/Hyperactivity Disorder Symptoms and Normal Behavior; Skew = skewness; Kurt = kurtosis; Rel = maximal reliability for the latent factors from Model 2.

[a] Raw accuracy score;

[b] Response time (in milliseconds for Stop Signal Test median reaction time, and in seconds for D- KEFS Color Word Interference Inhibition Switching and Trailmaking Letter-Number Switching);

[c] W-score;

[d] d-prime;

Table 2

Model fit indices.

Model	AIC	BIC	ABIC	χ^2	df	RMSEA [CI]	CFI	TLI	SRMR	χ^2 diff[a]	χ^2 diff[b]
Preliminary Hypothesized CFA Models											
Model 1a - 8 factors	45582.57	46170.39	45776.60	2203.37	436	0.069 [.066, .072]	0.80	0.77	0.08		
Model 1b - 7 factors	42251.91	42735.44	42411.52	851.56	303	0.046 [.043, .050]	0.90	0.89	0.08		
Correlated Hypothesized CFA Factors											
Model 2a - 8 factors	42133.71	42650.43	42304.28	742.42	296	0.042 [.038, .046]	0.92	0.91	0.06		
Model 2b - 7 factors	39423.83	39907.36	39583.44	748.57	303	0.042 [.038, .045]	0.92	0.91	0.06	6.46[c]	
Contrastive CFA Models											
Model 3a - 1 factor	43954.46	44338.44	44081.21	2279.01	324	0.080 [.081, .088]	0.66	0.63	0.13	1526.31**	1530.44**
Model 3b - 2 factors (P,R)	43060.74	43454.21	43190.62	1536.48	322	0.067 [.063, .070]	0.79	0.77	0.09	897.14**	955.36**
Model 3c - 3 factors (P,S,T)	42382.97	42785.92	42515.98	980.42	320	0.049 [.046, .053]	0.89	0.87	0.08	239.98**	237.28**
Model 3d - 3 factors (PC,PP,R)	42992.36	43390.56	43123.80	1488.85	321	0.066 [.062, .069]	0.80	0.78	0.09	663.44**	631.51**
Model 3e - 4 factors (PC,PP,S,T)	42312.60	42725.03	42448.74	925.23	318	0.048 [.044, .051]	0.89	0.88	0.08	157.84**	140.32**
Model 3f - 3 factors (PVe, Pvi, R)	43013.03	43411.24	43144.48	1502.02	321	0.066 [.063, .069]	0.79	0.78	0.09	737.13**	728.72**
Model 3g - 4 factors (PVe, PVi, S, T)	42329.95	42742.38	42466.09	937.55	318	0.048 [.044, .052]	0.89	0.88	0.09	176.96**	162.55**
Bifactor Model											
Model 4 - 5 specific factors	39301.57	39785.11	39461.19	649.31	303	0.037 [.033, .041]	0.94	0.93	0.07	−62.13**	n/a[d]

Note. AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; ABIC = Adjusted Bayesian Information Criterion; CFI = Comparative Fit Index; Chi sq. = Chi square; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual. TLI = Tucker Lewis Index. P = any performance measure; PC = computer performance; PP = paper performance; R = any report rating; S = self-report; T = teacher report; PVe = verbal performance; PVi = visual performance.

**
p < 0.001;

[a]Satorra-Bentler scaled chi-square difference test versus Model 2a; Model 2a produced a non-positive definite covariance matrix and fit details are provided only for completeness.

[b]Satorra–Bentler scaled chi-square difference test versus Model 2b

[c]Satorra-Bentler scaled chi-square difference test for Model 2a vs. Model 2b

[d]The Satorra-Bentler scaled chi-square difference test could not be computed for Model 4 versus Model 2b because both models estimated the same number of parameters (*df* = 303).

Table 3

Correlations among factors for the CFA models with 8 factors (Model 2a; *below* diagonal) and 7 factors (Model 2b; *above* diagonal).

	WM-S/M	WMU	INHIBIT	SHIFT	PLAN[a]	FLUENCY	SRL	MCOG
WMU	0.17	–	0.63	0.25	0.26	0.07	–0.03	0.23
INHIBIT	0.71	0.65	–	0.62	0.55	0.25	0.03	0.30
SHIFT	0.92	0.25	0.63	–	0.87	0.80	–0.06	0.37
PLAN[a]	0.96	0.29	0.46	0.81	–	0.64	–0.19	0.43
FLUENCY	0.66	0.06	0.25	0.80	0.61	–	–0.10	0.17
SRL	–0.22	–0.03	0.03	–0.07	–0.17	–0.10	–	0.12
MCOG	0.42	0.23	0.31	0.37	0.43	0.17	0.12	–

Note. Factor names: WM-S/M = Working Memory as Span/Manipulate; WMU= Working Memory as Updating; INHIBIT = Inhibition; SHIFT = Shifting; PLAN = Planning; SRL = Self-Regulated Learning; MCOG = Metacognition.

[a] As column header, PLAN represents the Working Memory as Span/Manipulate and Planning factor for Model 2b; as a row header, PLAN represents only the original Planning indicators. r's < |.16| not significant; r's > |.16| significant at 0.05.

Table 4

Standardized solution for Models 2a, 2b and 4.

	CFA (8 factors) Model 2a	CFA (7 factors) Model 2b	Bifactor (with 5 specific factors) Model 4	
	Loading (SE)	**Loading (SE)**	**Loadings (SE)**	**EF Loadings (SE)**
Shifting				
DKDFS	0.49 (0.06)	0.52 (0.05)		0.45 (0.06)
DKCS	0.52 (0.05)	0.55 (0.05)		0.54 (0.05)
DKIS	0.73 (0.05)	0.48 (0.06)		0.71 (0.04)
DKTS	0.55 (0.05)	0.73 (0.05)		0.58 (0.05)
WM-S/M + Plan				
WMLR	0.60 (0.10)	0.61 (0.05)	0.02 (0.10)	0.60 (0.04)
CBCK	0.60 (0.10)	0.61 (0.05)	0.27 (0.11)	0.49 (0.06)
TOL[a]	0.45 (0.05)	0.43 (0.04)	0.54 (0.15)	0.34 (0.04)
TOLEX[a]	0.52 (0.05)	0.49 (0.04)	0.34 (0.10)	0.41 (0.04)
PLAN[a]	0.51 (0.04)	0.51 (0.04)	0.09 (0.08)	0.48 (0.04)
WMU				
NBL1	0.78 (0.07)	0.78 (0.08)	0.73 (0.09)	0.22 (0.08)
NBL2	0.58 (0.08)	0.58 (0.08)	0.57 (0.09)	0.19 (0.07)
NBS1	0.63 (0.08)	0.64 (0.09)	0.57 (0.12)	0.15 (0.06)
NBS2	0.52 (0.09)	0.51 (0.09)	0.57 (0.12)	0.05 (0.06)
Inhibition				
GNG	0.53 (0.09)	0.36 (0.08)		0.18 (0.06)
SST	0.53 (0.09)	0.55 (0.09)		0.34 (0.04)
Fluency				
DKDF	0.51 (0.04)	0.51 (0.04)	0.14 (0.05)	0.48 (0.04)
DKCF	0.69 (0.03)	0.69 (0.03)	0.72 (0.21)	0.46 (0.04)
DKLF	0.69 (0.04)	0.69 (0.04)	0.38 (0.12)	0.53 (0.04)
Self-Regulated Learning				
EFEFF	0.71 (0.04)	0.71 (0.04)	0.75 (0.03)	0.18 (0.05)
SKPRF	0.58 (0.04)	0.59 (0.04)	0.72 (0.21)	−0.06 (0.05)
STRAT	0.90 (0.04)	0.90 (0.04)	0.38 (0.12)	−0.18 (0.04)
Metacognition				
BTI	0.93 (0.01)	0.93 (0.01)	0.84 (0.02)	0.42 (0.05)
BTMO	0.88 (0.01)	0.88 (0.01)	0.84 (0.02)	0.28 (0.05)
BTPO	0.96 (0.01)	0.96 (0.01)	0.89 (0.02)	0.37 (0.05)
BTOM	0.80 (0.02)	0.80 (0.02)	0.77 (0.02)	0.23 (0.05)
BTWM	0.95 (0.01)	0.95 (0.01)	0.88 (0.02)	0.36 (0.05)
SWAN	0.81 (0.02)	0.81 (0.01)	0.71 (0.03)	0.40 (0.04)

Note. DKDFS = D-KEFS Design Fluency Switching; DKCS = D-KEFS Verbal Fluency Category Switching; DKIS = D-KEFS Color Word Interference Inhibition Switching; DKTS = D-KEFS Trailmaking Test Letter-Number Switching; WMLR = WMTB-C Listening Recall; CBCK = Corsi Blocks Backwards Recall; TOL = Tower of London Accuracy; TOLEX = Tower of London Extra Moves; PLAN = WJ-III Planning; NBL1 = n-back 1 Letter; NBL2 = n-back 2 Letter; NBS1 = n-back 1Shape; NBS2 = n-back 2 Shape; GNG = Go No-Go; SST = Stop Signal Task; DKDF =

D-KEFS Design Fluency Filled+Empty Dots; DKCF = D-KEFS Verbal Fluency Categories; DKLF = D-KEFS Verbal Fluency Letters; EFEFF = CLS Effort/Self-Efficacy; SKPRF = CLS Skill/Preference; STRAT = CLS Learning Strategies; BTI = BRIEF Initiate; BTMO = BRIEF Monitor; BTPO = BRIEF Planning/Organization; BTOM = BRIEF Organization of Materials; BTWM = BRIEF Working Memory; SWIN = SWAN Inattention. WM-S/M = Working Memory as Span/Manipulate; WMU = Working Memory as Updating.

[a]These variables were indicators of the planning factor in Model 1.