Towards a Multivariate Assessment of Executive Functions

by

Justin Elliott Karr
Master of Science, University of Victoria, 2013
Bachelor of Science, Western Oregon University, 2011

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Psychology

Supervisory Committee

Towards a Multivariate Assessment of Executive Functions

by

Justin Elliott Karr
Master of Science, University of Victoria, 2013
Bachelor of Science, Western Oregon University, 2011

Supervisory Committee

Mauricio A. Garcia-Barrera, Ph.D., Department of Psychology
Supervisor

Scott M. Hofer, Ph.D., Department of Psychology
Departmental Member

Grant L. Iverson, Ph. D., Department of Physical Medicine & Rehabilitation, Harvard Medical School
Outside Member

Abstract

Supervisory Committee
Mauricio A. Garcia-Barrera, Ph.D., Department of Psychology
Supervisor
Scott M. Hofer, Ph.D., Department of Psychology
Departmental Member
Grant L. Iverson, Ph.D., Department of Physical Medicine & Rehabilitation, Harvard Medical School
Outside Member

Objective: This work consisted of three research projects bridged by their focus on a multivariate assessment of executive functions in research and practice: (a) a systematic review and re-analysis of latent variable studies on executive function test batteries, (b) a confirmatory factor analysis (CFA) of the Delis-Kaplan Executive Function System (D-KEFS), the most commonly administered executive function test battery in clinical practice, and (c) the derivation of multivariate base rates for the D-KEFS, offering a psychometric resource with direct applications to clinical practice. **Method:** ***Systematic review.*** The systematic review identified 45 eligible samples (*N*=9,498 participants, mean age range: 3.01-74.40 years-old) and 21 correlation matrices eligible for re-analysis, comparing seven competing models including the most commonly evaluated factors: updating/working memory, inhibition, and shifting. Model results were summarized based on the mean percent accepted (i.e., mean rate at which models both properly converged and met fit thresholds: CFI≥.90/RMSEA≤.08). ***CFA.*** Using adults from the D-KEFS normative sample (*N*=425; 20-49 years-old), eight alternative measurement models were evaluated for a subset of D-KEFS tests. Factors from the accepted measurement model predicted three tests measuring constructs less often evaluated in the executive function literature: abstraction, reasoning, and problem solving. ***Base rates.*** The frequency of low scores occurring among the D-KEFS normative sample (*N*=1,050; 16-89 years-old) was calculated for the full D-KEFS and two brief batteries using stratifications for age, education, and intelligence. **Results:** ***Systematic review.*** The most often accepted models varied by age (preschool=one/two-factor; school-age=two/three-factor; adolescent/adult=three/nested-factor; older adult=two/three-factor), and most frequently included updating/working memory, inhibition, and shifting factors. The nested-factor and three-factor models were accepted most often and at similar rates among adult samples: 33-34% and 25-32%, respectively. No model was accepted most often for child/adolescent samples, but those with shifting differentiated garnered less support. ***CFA.*** A three-factor model including inhibition, shifting, and fluency fit the data well (CFI=0.938; RMSEA=0.047), although a two-factor model merging shifting/fluency fit similarly well (CFI=0.929; RMSEA=0.048). A bifactor model fit best (CFI=0.977; RMSEA=0.032), but rarely converged. Shifting best predicted tests of reasoning, abstraction, and problem solving (p<0.05; R^2=0.246-0.408). ***Base rates*****.** Low scores, based on commonly used clinical cutoffs, occurred frequently among healthy adults. For a three-test, four-test, and full D-KEFS battery, 62.8%, 71.8%, and 82.6% obtained ≥1 score(s) ≤16th percentile, respectively, and 36.1%, 42.0%, 50.7%, obtained ≥1 score(s) ≤5th percentile, respectively. The frequency of low scores increased with lower intelligence and fewer years of education. **Discussion:** The systematic review effort did not identify a definitive model of executive functions for either adults or children/adolescents, demonstrating the continued need to re-evaluate the conceptualization and measurement of this construct in future research. The D-KEFS CFA offers some evidence of clinical measures capturing theoretical constructs, but is not directly translatable into clinical practice; while the multivariate base rates are useful to clinicians, but do not bridge theory and assessment. This research reaffirms the elusive nature of executive functions in both research and clinical spheres, and represents a step forward in an enduring scientific process *towards* a true understanding of this mysterious construct.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to acknowledge my colleagues, Mr. Corson N. Areshenkoff and Dr. Philippe Rast, for their support and consultation; and of course, for sharing their statistical wisdom with me. I would also like to thank Robyn E. Kilshaw and Ryan J. Tonkin for the many hours that they volunteered to assist me with data extraction for the systematic review.

Dedication

I dedicate this dissertation to my mother and father. All that I accomplish is because of you.

Prologue

This dissertation is composed of three chapters all focused on multivariate approaches to the assessment of executive functions. The observation that ultimately inspired this dissertation came through my training as a scientist-practitioner. I noticed substantial differences between research and clinical practices in how neuropsychologists evaluated executive functions, which led to two broad research questions: first, what does the multivariate research on executive functions suggest about the structure of this construct; and second, how can clinicians begin to implement multivariate approaches to executive function assessment in clinical practice?

The first question was addressed in the first chapter, which involved a systematic review and re-analysis of latent variable studies evaluating the structure of executive functions through confirmatory factor analyses. The second question was addressed in the second and third chapters, which both made use of the normative data from the Delis-Kaplan Executive Function System (D-KEFS), the most commonly administered executive function test battery in neuropsychological practice. The second chapter involved a confirmatory factor analysis of the D-KEFS to identify whether its latent structure aligned with previous research; while the third chapter described the derivation of multivariate base rates, quantifying the normal prevalence of low scores on the D-KEFS among healthy adults.

The three chapters are inter-related in that their findings inform each other, but they are largely standalone research contributions, each with specific aims, methods, conclusions, and limitations. However, as reflected in the dissertation's title, they share the common orientation of moving *Towards a Multivariate Assessment of Executive*

Functions in both scientific and applied settings. That orientation aligns with the observation that originally inspired all three chapters, with the overall aim of bridging research and practice.

Chapter 1: The Unity and Diversity of Executive Functions: A Systematic Review and Re-Analysis of Latent Variable Studies

Abstract

Confirmatory factor analysis (CFA) is frequently applied to the measurement of executive functions, since first used to identify a three-factor model of inhibition, updating, and shifting; however, subsequent CFAs have supported inconsistent models across the lifespan, ranging from unidimensional to nested-factor models (i.e., bifactor without inhibition). This systematic review aimed to summarize CFAs on performance-based tests of executive functions and determine best-fitting models by reanalyzing summary data. Eligible CFAs included 9,498 participants across 45 samples ($\bar{x}$ age range: 3.01 to 74.40). The most frequently accepted models varied by age (i.e., preschool=one/two-factor; school-age=two/three-factor; adolescent/adult=three/nested-factor; older adult=two/three-factor), and most often included updating/working memory, inhibition, and shifting factors. A bootstrap re-analysis simulated 5,000 samples from 21 correlation matrices (i.e., 11 child/adolescent; 10 adult) with indicators for the three most frequent factors, fitting seven competing models. Model results were summarized based on the mean percent accepted: the average rate across studies at which models both properly converged and met fit thresholds (i.e., CFI≥.90/RMSEA≤.08). No model consistently converged and met fit criteria in all samples. Among adult samples, the nested-factor and three-factor models were accepted most often and at similar rates: 33-34% and 25-32%, respectively. Among child/adolescent samples, no model was accepted most often, but those with shifting differentiated garnered less support. Results suggested increased differentiation of executive function with age, indicating a one/two-factor model for child/adolescent samples and a three/nested-factor model among adults.

However, low rates of model acceptance suggest possible bias towards the publication of well-fitting, but potentially non-replicable models with underpowered samples.

Introduction

In the past decade, executive functions have garnered a significant amount of clinical and research attention in regards to their definition and measurement (Barkley, 2012; Chan, Shum, Toulopoulou, & Chen, 2008; Jurado & Rosselli, 2007; Pickens, Ostwald, Murphy-Pace, & Bergstrom, 2010). There has also been considerable interest in their predictive validity for clinical and societal outcomes (e.g., childhood problem behaviors; Espy et al., 2011; instrumental activities of daily living; Cahn-Weiner, Boyle, & Malloy, 2002; Bell-McGinty, Podell, Franzen, Baird, & Williams, 2002; personal finances, health, criminality, substance dependence; Moffitt et al., 2011). However, despite a large body of research on executive functions, the field lacks both a universal definition and an agreed upon form of measurement (Barkley, 2012; Baggetta & Alexander, 2016). Throughout the history of neuropsychology, executive functions have received diverse definitions. Before the term 'executive functions' debuted in the neuropsychological literature (Lezak, 1982), researchers had linked the term 'executive' with both frontal lobe functioning (Pribram, 1973) and control over lower-level cognitive abilities (Baddeley & Hitch, 1974).

Early models of executive functions detailed a 'central executive' that managed lower-level cognitive processes in the context of working memory (Baddeley & Hitch, 1974), while other researchers extended this concept to a system of conscious control over attention (i.e., the Supervisory Attentional System [SAS]; Norman & Shallice, 1986). Based on clinical conceptualizations of frontal processes (Luria, 1966), the functions of the SAS were also attributed to the frontal lobes. These early researchers painted a relatively unitary picture of frontal functioning and executive functions –

although they did yet not use this term – where a localized neural substrate underlies a single control function. However, successive definitions of executive functions have demonstrated the diversity of abilities falling under this umbrella term (Barkley, 2012; Baggetta & Alexander, 2016); and, further, an established body of neuropsychological research has implicated multiple brain regions that interact with the frontal lobes (e.g., parietal lobes, cerebellum) in the expression of executive functions (Alvarez & Emory, 2006; Collette, Hogge, Salmon, & Van der Linden, 2006; Keren-Happuch, Chen, Ho, & Desmond, 2014).

Prior to unitary models of higher-order cognition, clinicians commonly evaluated many of the abilities now considered executive functions (e.g., planning, self-regulation, fluency) long before scholars clustered these abilities into a common construct (Lezak, 1976). The debate between the unity and diversity of frontal functioning (Teuber, 1972) and executive functions (Miyake et al., 2000) has perpetuated for decades, although early definitions for executive functions (e.g., Lezak, 1983; Welsh & Pennington, 1988), and nearly all definitions that followed (Barkley, 2012; Baggetta & Alexander, 2016; Jurado & Rosselli, 2007), have described the construct as multidimensional. The earliest definition of executive functions described the construct as having "four components" (Lezak, 1983, p. 507), with sequential descriptions defining executive functions as an "umbrella term" (Chan et al., 2008, p. 201) for a family of "poorly defined" (Burgess, 2004, p. 79), "meta-cognitive" (Oosterlaan, Scheres, & Sergeant, 2005, p. 69) or "cognitive control" (Friedman et al., 2007, p. 893) processes "used in self-regulation" (Barkley, 2001, p. 5).

Roughly 20 years ago, researchers had proposed some 33 definitions for executive functions (Eslinger, 1996). The labels and tests for executive functions have been so diverse within the published research that one recent literature review identified 68 sub-components of executive function, reduced to 18 sub-components following an analysis that removed semantic and psychometric overlap between terms (Packwood, Hodgetts, & Tremblay, 2011). The authors of this review conceded that the large number of executive functions posited by various researchers lacked parsimony. In turn, despite years of research on diverse executive functions, the exact number of constructs rightfully labeled executive functions remains largely unknown.

Understanding the number of executive functions supported by the neuropsychological literature first requires an understanding of their measurement. The traditional measurement of executive functions in both research and clinical practice has relied largely on the use of single tests (Baggetta & Alexander, 2016; Chan et al., 2008; Rabin, Barr, & Burton, 2005; Rabin, Paolillo, & Barr, 2016). Tests purported to measure executive functions have varied significantly across studies, with task characteristics sometimes having a greater effect on test performances than the personal and diagnostic features of participants (e.g., age, gender, nature of reading difficulties; Booth, Boyle, & Kelly, 2010). With the heterogeneity of available tests of executive functions, researchers likely inferred that the many tests used to measure executive functions did not all necessarily measure the same unitary construct; however, this inference has resulted in the over-naming of task-specific behaviors as separable executive sub-components (Packwood et al., 2011). This approach lacks discretion and ignores the high interrelatedness between both neuropsychological tests and the terms used to describe their outcomes.

A rich history of published research has explored the correlations between tests of executive functions using a factor analytic approach (Royall et al., 2002). The first factor analyses on executive functions used an exploratory approach that did not impose any hypothesized correlational structure on the battery of tests. The first appearance of an executive function measure in a factor analysis observed the Stroop test loading on a factor involved in the cognitive control over attention (Barroso, 1983). Sequential studies found a heterogeneous number of factors, ranging from a minimum of one factor (e.g., Deckel & Hesselbrock, 1996; Della Sala, Gray, Spinnler, & Trivelli, 1998) to as many as six factors (Testa, Bennett, & Ponsford, 2012). In multiple contexts, the outcomes of many tasks measuring executive functions loaded together on task-specific factors rather than loading onto common factors composed of indicators from multiple tests (e.g., Cirino, Chapieski, & Massman, 2000; Grodzinsky & Diamond, 1992; Levin et al., 1996; Latzman & Markon, 2010). These findings suggest that the indicators included in these exploratory analyses grouped based on common method variance rather than underlying executive constructs (Barkley, 2012). These task-specific factors may derive largely from the statistical limitations of an exploratory approach, where the relationships between tasks lack a hypothesized structure and potentially group together due to non-executive abilities that also contribute to task performance (Hughes & Graham, 2002).

Many of the tasks employed to measure executive functions have an underlying multidimensional structure (e.g., the Wisconsin Card Sorting Test, Greve et al., 2005; the Trail Making Test, Sanchez-Cubillo et al., 2009), with many different cognitive abilities interacting to explain a given performance (Duggan & Garcia-Barrera, 2015). Executive function tests have a reputation for task impurity, whereby many non-executive abilities explain performances on tests purported to measure executive functions (Burgess, 1997; Miyake & Friedman, 2012; Phillips, 1997). As a rule, neuropsychological tests do not provide a pure measurement of a specific cognitive domain and researchers do not assert that tests have impeccable construct validity. Nonetheless, the use of a single test as an indicator of executive functions ignores the impact of task impurity on neuropsychological outcomes (Baggetta & Alexander, 2016).

To combat task impurity, a seminal article in the research on executive functions (i.e., Miyake et al., 2000) used a confirmatory factor analysis to assess the relationship between interrelated manifest variables commonly used in cognitive research as measures of three executive functions: the “shifting of mental sets, monitoring and updating of working memory representations, and inhibition of prepotent responses” (p. 50). These researchers constructed a battery of diverse tasks that tapped into three established executive functions, selected based on a rich history of research. They assigned these tasks to hypothesized factors based on their common construct variance and found that a three-factor model best fit the data. In turn, they demonstrated the promise of confirmatory factor analysis at providing purer estimates of executive functions, not contaminated by non-executive method variance. Following this approach, updating, inhibiting, and shifting have all garnered further support through a series of subsequent empirical studies reporting similar three-factor solutions from confirmatory factor models of cognitive tasks (e.g., Friedman et al., 2006, 2008; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Vaughan & Giovanello, 2010).

The published research on measurement models for executive functions has burgeoned in the new millennium (Willoughby, Holochwost, Blanton, & Blair, 2014). The solutions from confirmatory factor analyses accepted by past researchers have varied significantly in terms of the number of factors identified, ranging from a single factor at early age (e.g., Brydges, Reid, Fox, & Anderson, 2012; Hughes, Ensor, Wilson, & Graham, 2010; Wiebe, Espy, & Charak, 2008) and late age (e.g., de Frias, Dixon, & Strauss, 2006; Ettenhofer, Hambrick, & Abeles, 2006) to as many as five at young adulthood (i.e., Fournier-Vicente, Larigauderie, & Gaonac'h, 2008). The first measurement model reported for executive functions remains the most popularly discussed in the literature (Miyake et al., 2000); however, they do not necessarily represent the full gamut of empirically supported executive functions (Jurado & Rosselli, 2007) and Miyake and colleagues (2000) never described them as an exhaustive list of executive functions. The terms most commonly used to label executive functions include planning, working memory, fluency, inhibition, and set-shifting (Packwood et al., 2011); however, these terms simply present most frequently in the literature, and do not necessarily represent a comprehensive list of relevant executive functions (Barkley, 2012).

The discussion of how many executive functions exist implies that the many abilities labeled "executive" represent separable cognitive capacities; however, each factor does not necessarily represent an orthogonal construct, considering the high correlations often observed between the latent variables of different functions (e.g., .63 to .65, Lehto et al., 2003; .42 to .63, Miyake et al., 2000; .68 to .81, Vaughan & Giovanello, 2010). Working memory capacity and vocabulary both significantly predict outcomes on fluency tasks (Unsworth, Spillers, & Brewer, 2011) and may represent an outcome of working memory interacting with the lexicon (Shao, Janse, Visser, & Meyer, 2014). Similarly, planning represents a higher-order construct, with updating, shifting, and inhibition potentially operating in a collaborative fashion to explain performances on planning-related tasks (Miyake & Friedman, 2012). The exact relationship between updating, shifting, and inhibition is still not defined, as more recent studies have found that the majority of variance in these three executive functions may be explained by a common higher-order dimension (e.g., Fleming, Heintzelman, & Bartholow, 2016; Friedman et al., 2008; Ito et al., 2015).

Considering the conceptual and empirical overlap between updating, shifting and inhibition, researchers have begun re-evaluating the shared variance between the constructs through an alternative measurement model (e.g., Friedman et al., 2008, 2016; Friedman, Corley, Hewitt, & Wright, 2009; Friedman, Miyake, Robinson, & Hewitt, 2011). Using a nested factor model in repeated analyses of the same dataset, Friedman and colleagues (2008, 2009, 2011, 2016) had all indicators load on a general factor and indicators for updating and shifting co-load on factors specific to those constructs. Because the general factor fully explained the variance in inhibition, the researchers did not include it as a specific factor, with its indicators loading only on the general factor. This model represents an incomplete bifactor model (Chen, West, & Sousa, 2006) and demonstrates a substantial amount of shared variance between indicators across factors in a multidimensional test battery. These findings emphasize the need to consider both general and specific dimensions when explaining performances on test batteries evaluating executive functions.

Considering the recent conclusions of Miyake and Friedman (2012) and the many published confirmatory factor analyses supporting multidimensional solutions using performance-based tests (Willoughby et al., 2014), the latent variable research on executive functions has reached a point of requiring both knowledge synthesis and a re-evaluation of previously supported factor solutions. Foremost, the published literature on executive function measurement models has never been comprehensively summarized, and a systematic review would identify the factor models with the most empirical support. Further, few researchers aside from Friedman and colleagues (2008, 2009, 2011, 2016) have evaluated the presence of a common executive function dimension through a nested factor modeling approach (e.g., Fleming et al., 2016; Garza et al., 2014; Ito et al., 2015; Kramer et al., 2014), but all of these researchers have found a robust general factor. In turn, those researchers not exploring a general dimension potentially over-estimate the diversity of executive function factors, and a re-analysis of previous findings could evaluate whether or not a nested factor model offers superior statistical fit to a multidimensional solution.

The current study aimed to (a) determine the empirical support for measurement models of executive functions proposed by past researchers, (b) identify the number of purported executive functions supported by confirmatory factor analyses in the current literature, and (c) determine which published measurement model best fits summary data across studies. To fulfill the first two aims, the current study involved a broad systematic review of research reporting confirmatory factor analyses on batteries of performance-based tasks evaluating executive functions, summarizing both the frequency of model solutions (e.g., unidimensional, three-factor, nested factor models) and the rate at which different factors were included in accepted measurement models (e.g., inhibition, updating, shifting, etc.). Considering the significant heterogeneity between the measurement models evaluated by past researchers, the approach to the third aim required a more precise focus on comparable studies, and ultimately considered only those studies assessing the most frequently evaluated factor model within the published literature (i.e., the three-factor measurement model of inhibition, shifting, and updating/working memory; Miyake et al., 2000). The results of these comparable studies were re-analyzed and fitted to competing factor solutions based on the published literature. By fulfilling these aims, the current review described the diversity of existing latent variable research on executive functions and further clarified the level of empirical evidence behind the most common factor solutions proposed by past researchers.

Method

The report of this systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (Moher, Liberati, Tetzlaff, Altman, and the PRISMA Group, 2009). Prior to the literature search, inclusion criteria were established to identify appropriate articles. For inclusion, articles needed to (a) involve a sample or sub-sample of cognitively healthy participants (i.e., without a neurodevelopmental or neurological disorder known to significantly impact cognitive performance) and (b) report a confirmatory factor analysis of a multidimensional measurement model of executive function. Following this criterion, studies that included multiple factors that could be conceptualized as executive functions, but not directly specified as dimensions of executive function or a synonymous construct (e.g., executive control) by the authors were ineligible (e.g., Unsworth, Spillers, & Brewer, 2011; McVay

& Kane, 2012). As well, measurement models of solely sub-components of executive function were ineligible (e.g., inhibition, Aichert et al., 2012, Friedman & Miyake, 2004; effortful control, Allan & Lonigan, 2011, 2014; problem solving; Cinan, Özen, & Hampshire, 2013; Scherer & Tiemann, 2014). Eligible models needed to include (c) a minimum of two indicators, deriving from separate tests, per construct evaluated and (d) only performance-based cognitive or neuropsychological outcomes as indicators for the executive function factor(s) (i.e., studies including biometrics, rating scales, or symptom inventories as indicators were ineligible), deriving from (e) at least three separate cognitive or neuropsychological tests (i.e., measurement models evaluating the factor structure of multiple outcomes from a single neuropsychological test were ineligible). Lastly, the articles needed to (f) be published in either a peer-reviewed journal or academic book and (g) be written in the English language. For inclusion in the re-analysis, which synthesized a comparable sub-sample of studies testing the most commonly evaluated measurement model in the literature, the articles needed to meet all aforementioned criteria, but also had to have (h) evaluated a measurement model including factors of inhibition, shifting, and updating (or analogous constructs; e.g., mental set-shifting, switching, working memory, etc.) and (i) provide sufficient summary data for re-analysis (i.e., at least a correlation matrix for all tests included in the model).

Literature Search

The systematic literature search occurred in August 2016 and involved online searches of the following databases, with search restrictions in parentheses: PsycInfo (Publication type – Peer-reviewed journals, All books; Methodology – Empirical studies, Quantitative studies; Population group – Human; Language – English), PsycArticles

(Publication type – Empirical studies, Quantitative studies; Population group – Human), MedLine (Publication type – Journal article; Population group – Human; Language – English), and CINAHL (Publication type – Journal article, Book, Book chapter, Research, Statistics; Language – English). Search results were restricted to literature published from 1998 to the time of the electronic search, with this date range selected to capture articles following the publication of Miyake et al. (2000) and any articles published just prior to this study that may have involved a confirmatory factor analysis of tests of executive functions. The search protocol involved the following Medical Subject Headings (MeSH), Psychological Index Terms (Tuleya, 2009), and search terms:

> ((MM "Factor Analysis" OR MM "Factor Structure" OR MM "Goodness of Fit" OR MM "Structural Equation Modeling") OR (MM "Factor Analysis, Statistical" OR MM "Models, Statistical") OR (“confirmatory factor analysis” OR “CFA” OR "latent variable")) AND ((DE "Executive Function" OR DE "Cognitive Control" OR DE "Set Shifting" OR DE "Task Switching" OR MM "self regulation") OR (MM "Executive Function" OR MM "Inhibition (Psychology)" OR MM "Problem Solving") OR ("executive function*" OR "self-regulat*"))

All retrieved search results were screened twice to ensure that no study went overlooked (Edwards et al., 2002). Following the electronic search, reference lists from peer-reviewed journals were manually searched over the course of data extraction and manuscript preparation, identifying any articles missed by the electronic search protocol (See Figure 1, for a flow diagram of the systematic review process along with the number of articles identified). A reference list of full-text articles reviewed during the literature search, but ultimately not included in the systematic review, is provided in Appendix A, organized by their reason for exclusion.

Data Extraction

Two independent reviewers extracted relevant information from each article through use of a common data collection spreadsheet. Both reviewers extracted variables related to study characteristics (i.e., authorship, year of publication), sample characteristics (i.e., percent female, mean age, mean years of education, ethnic composition), model characteristics (i.e., name of dependent variables and respective factors), and factor analytic results for accepted measurement models (i.e., χ^2 value and respective *p*-value; comparative fit index, CFI; root mean squared error of approximation, RMSEA). For samples eligible for the re-analysis, summary data necessary for a re-analysis of the measurement model was also extracted (i.e., sample size, means/standard deviations, correlation/covariance matrix).

To quantify study quality, reviewers rated articles based on a scale developed specifically for the current review. The majority of confirmatory factor analytic studies involve observational research designs with one time point of data collection (Willoughby et al., 2014), which represents one of the lowest levels of scientific evidence (OCEBM Levels of Evidence Working Group, 2011). Few instruments for rating the quality of this level of research exist in the current literature (Sanderson, Tatt, & Higgins, 2007; Vandenbroucke et al., 2007). In turn, the current systematic review strategy applied eleven criteria to rate study quality. These criteria were based largely on standard publication practices for factor analyses (Schreiber, Nora, Stage, Barlow, & King, 2006), with each item scored as either met (1 point) or not met (0 points) and summed for a total study quality score (range: 0-11). The study quality rating scale included the following items:

> (1) the researchers reported a sample size with $\hat{\pi} \geq .80$ to reject the null hypothesis (RMSEA $\geq .05$) for a model obtaining a perfect RMSEA (Hancock, 2006), (2) listed at least two demographic variables for each sample evaluated (e.g., mean age, gender composition), (3) indicated that data screening/cleaning for outliers or data transformations to ensure normality was conducted, (4) provided a path diagram of at least one measurement model evaluated or a structural model including all variables from the accepted measurement model, (5) reported the results of a χ^2 goodness-of-fit test and at least two alternative fit indices (e.g., RMSEA, CFI, etc.), (6) listed all of the loadings and (7) residuals for at least one measurement model or structural model evaluated, (8) provided inter-factor correlations for at least one of the multidimensional measurement models or structural models evaluated (if constrained to zero, the authors reported this constraint in the manuscript), (9) reported the means and standard deviations for all manifest variables included in the measurement model, (10) provided a correlation or covariance matrix including all manifest variables included in the measurement model, and (11) had at least three indicators loading on each latent factor in every measurement model evaluated (Roberts & Grover, 2009).

The selection of the power criterion in this scale was based on post-hoc power analyses for model fit. A power ($\hat{\pi}$) cutoff of $\geq .80$ was selected as a conventional threshold in power analysis (Cohen, 1992). Hancock (2006) provided tables to calculate post-hoc power to reject the null hypothesis (i.e., RMSEA $\geq .05$) based on three RMSEA values

(.00, .02, .04). The tables for the perfect RMSEA value (i.e., .00) were used to determine whether models met sufficient power (i.e., $\hat{\pi} \geq .80$) because (a) many studies reported perfect RMSEA values and (b) these tables listed the smallest required sample sizes to meet this threshold. Stricter thresholds would have resulted in few or no studies meeting this criterion.

Re-Analysis

All articles eligible for the re-analysis provided a correlation matrix for their test battery and tested the same three-factor model, including factors of inhibition, updating, and shifting or analogous constructs. One study included in the re-analysis (Hedden & Yoon, 2006) reported two factors that could be considered inhibition-related factors (i.e., prepotent response inhibition and resistance to proactive interference). Because prepotent response inhibition was most analogous to the inhibition factor included in other measurement models also eligible for the re-analysis, this factor was included as the inhibition factor in all models run using the correlation matrix for this study, while the resistance to proactive interference factor was left out.

The re-analysis involved two primary aims that rationalized the methodological approach. First, not all researchers examined all factor models supported by the literature with their dataset, and a re-analysis specifying multiple possible measurement models would determine if a specific factor model tended to fit best across published samples. Second, the risk for publication bias was of concern, because most publications identified in the systematic review reported small sample sizes and excellent-fitting models that converged without any errors.

The correlation matrix was re-analyzed by specifying seven alternative measurement models: a unidimensional model, three two-factor models that merged two of the first-order factors (i.e., inhibition = updating; updating = shifting; inhibition = shifting), a three-factor model (i.e., inhibition, updating, and shifting), a nested factor model (i.e., a common executive function bifactor, with shifting-specific and updating-specific factors co-loading on their select indicators and no inhibition-specific factor), and a bifactor model (i.e., a common executive function bifactor with specific factors for inhibition, shifting, and updating). See Figure 2 for a visual representation of each model. Five of these seven models (i.e., all but the bifactor model) were identified as published factor solutions by at least one study in the systematic review. While the full bifactor model was not accepted by any researchers, it was tested as a comparison point for the nested factor model (as done originally by Friedman et al., 2008), permitting evaluation of whether the removal of the inhibition-specific factor improved the fit of the model.

The re-analysis was conducted through a parametric bootstrap simulation based on the published correlation matrix where the data from each study were assumed to be multivariate normal with the observed correlation matrix considered equivalent to the population correlation matrix. For each sample, correlation matrices were computed for 5,000 simulated datasets of equal sample size to that of the original study. For all 5,000 correlation matrices, each factor model was fit to the data. Fit indices were calculated for models that “properly converged,” which means the model converged without any errors that would indicate a solution was inadmissible or the estimates were not trustworthy (e.g., a correlation above 1.0, negative residual variances, a non-positive definite latent variable covariance matrix). Throughout the rest of this manuscript, the terms properly converged and converged will be used synonymously. For all samples that properly converged, the CFI and RMSEA were calculated. All factor variances were fixed to 1.0 to set the metric for the factor, and all loadings were freely estimated for all models, with one exception: models with only two indicators on any specific factor in the bifactor or nested factor models had the loadings for those indicators set to be equal for purposes of model identification, as done by previous researchers (Canivez, 2014; Watkins, 2010). The bootstrap re-analysis was conducted in R (R Core Team, 2013), with all factor models fit using the Lavaan package (Rosseel, 2012). The full list of correlation matrices, syntax, and code are provided in Appendix B.

Bootstrapping method validation. The correlation matrices for the 16-69 year-old sample from the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV; *N* = 1,800; Wechsler, 2008) and the 6-16 year-old sample from the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V; *N* = 2,200; Wechsler, 2014) were re-analyzed using the bootstrapping method as a way of validating the approach. The WAIS-IV and WISC-V models were based on confirmatory factor analyses conducted using large, nationally stratified normative samples. The four-factor measurement model for the WAIS-IV has been replicated in a re-analysis (Weiss, Keith, Zhu, & Chen, 2013a) and the newly introduced five-factor model for the WISC-V has been previously postulated with older versions of the test battery (Weiss, Keith, Zhu, & Chen, 2013b). Although these models are not without controversy (Canivez & Kush, 2013), testing these models using the bootstrapping approach would determine whether a frequently evaluated and replicated model consistently produces model fit indices within an acceptable range.

The models specified for these correlation matrices were those reported for all primary and secondary subtests as the best fitting models in the technical manuals for each test. For the WAIS-IV, the model was a second-order factor model with four first-order factors (i.e., Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed) and 15 manifest variables, with a co-loading of Arithmetic on Verbal Comprehension and Working Memory and a co-loading of Figure Weights on Perceptual Reasoning and Working Memory. The errors for Digit Span and Letter-Number Sequencing were also allowed to correlate in this model. For the WISC-V, the model was a second-order factor model with five first-order factors (i.e., Verbal Comprehension, Visual Spatial, Fluid Reasoning, Working Memory, Processing Speed) and 16 manifest variables, including a constrained loading of 1.0 from Fluid Reasoning onto the second-order factor and a three-way co-loading of Arithmetic onto three first-order factors: Verbal Comprehension, Fluid Reasoning, and Working Memory.

Model Fit Interpretation. Model fit was evaluated by use of the CFI and RMSEA. These fit indices were selected for three reasons. First, these indices are commonly reported in the executive function literature, which is why they were included as extracted data elements for the systematic review. The majority of eligible studies reported these fit indices, and researchers within this field are familiar with their use. Second, they are not sensitive to sample size (Fan, Thompson, & Wang, 1999), which was important because the sample sizes varied substantially between studies. And third, they provide a common metric that is comparable across models and offer standard cutoff criteria to guide model selection and inference. Lenient and strict fit thresholds were used to guide model selection for both the CFI and RMSEA. For the CFI, the lenient and strict thresholds were ≥ 0.90 (Bentler & Bonett, 1980) and ≥ 0.95 (Hu & Bentler, 1999), respectively; and for the RMSEA, the lenient and strict thresholds were ≤.05 and ≤.08, respectively (Browne & Cudeck, 1993). The RMSEA was also a good choice because it favors parsimony (Hooper, Coughlan, & Mullen, 2008), which was meaningful when comparing models that ranged from simple unidimensional models to those with far more estimated parameters, such as the bifactor model.

The simulated data were interpreted based on the percent of models that properly converged and the percent of models that both converged and met lenient and strict cutoffs for the CFI and RMSEA. Across studies, the means and medians of these percentages were taken to identify the frequency at which a researcher with data from a battery of executive function tests would (a) have their proposed model converge without any errors that would affect inference and (b) meet standard fit criteria.

Results

Systematic Review

The literature review identified 39 articles meeting eligibility criteria for the systematic review reporting measurement models for 45 different samples. Among those eligible studies, 17 articles provided sufficient data for the re-analysis of 21 samples. A large set of studies examined for the current review pulled participants from the Victoria Longitudinal Study (de Frias, Dixon, & Strauss, 2006, 2009; McFall et al., 2013, 2014; Sapkota, Vergote, Westaway, Jhamandas, & Dixon, 2015; Thibeau, McFall, Wiebe, Anstey, & Dixon, 2016), the Colorado Longitudinal Twin Study (Friedman et al., 2006, 2007, 2008, 2009, 2011, 2016), and the Family Life Project study (Willoughby, Blair, & The Family Life Project Investigators, 2016; Willoughby, Blair, Wirth, Greenberg, & The

Family Life Project Investigators, 2010, 2012a; Willoughby, Wirth, Blair, & The Family Life Project Investigators, 2012b) with definitive or potential overlap among the participants included in their analyses. Some cross-sectional studies also reported analyses for the same participant data across different articles (Miller, Giesbrecht, Müller, McInerney, & Kerns, 2012; Miller, Müller, Giesbrecht, Carpendale, & Kerns, 2013; van der Ven et al., 2012, 2013; Usai, Viterbori, Traverso, & De Franchis, 2014; Viterbori, Usai, Traverso, & De Franchis, 2015; Rose, Feldman, & Jankowski, 2011, 2012). To avoid representing the same participants twice in the review, the studies involving the largest samples and the most executive function tasks were ultimately included in the systematic review and re-analysis (de Frias et al., 2009; Friedman et al., 2011; Miller et al., 2012; Rose et al., 2012; van der Ven et al., 2013, Willoughby et al., 2012a).

Most studies reporting confirmatory factor analyses on executive functions involved cross-sectional research designs; and for the limited amount of longitudinal studies identified, only one wave of measurement per study was represented in the current review and re-analysis. For one longitudinal study evaluating the same battery of executive function tasks at multiple time points, the data from the first wave were considered for the current review and re-analysis (i.e., de Frias et al., 2009). The consideration of just the first wave data made the study design more comparable to other studies in the review; however, in contexts where the task battery changed, the wave with the most available executive function tasks or the most complete summary data was considered in the current review (i.e., Willoughby et al., 2012a; Lee et al., 2013).

Qualitative Synthesis

Demographics of samples evaluated. Table 1 provides the demographic characteristics for each sample included in the systematic review along with an estimate of study quality. Among the samples reported by studies included in the systematic review, 9 samples (*n* = 2,614; $\bar{x}$ % female = 49.81%) consisted of preschool aged children ($\bar{x}$ age range: 3.01 to 5.77 years), 15 samples (*n* = 2,374; $\bar{x}$ % female = 48.54%) consisted of school-aged children ($\bar{x}$ age range: 6.42 to 11.88 years), 3 samples (*n* = 1,040; $\bar{x}$ % female = 48.87%) consisted of adolescents ($\bar{x}$ age range: 14.41 to 17.30 years), 8 samples (*n* = 1,812; $\bar{x}$ % female = 51.27%) consisted of adults ($\bar{x}$ age range: 19.75 to 25.70 years), and 8 samples (*n* = 1,112; $\bar{x}$ % female = 61.44%) consisted of older adults ($\bar{x}$ age range: 60.24 to 74.40 years). Two studies evaluated samples with participants spanning multiple age groups (*n* = 546), including a child to young adult sample ($\bar{x}$ age range: 7.20 to 20.80 years; Huizinga et al., 2006) and a merged young and older adult sample ($\bar{x}$ age range: 21.00 to 71.00 years; Pettigrew & Martin, 2014). Overall, 9,498 participants ($\bar{x}$ % female = 52.56%) were represented in the systematic review.

Among the 18 samples with some race or ethnicity information provided, 10 samples were predominantly White, 3 samples were majority non-White, and 5 samples were identified as ethnically Chinese (Lee et al., 2012; Xu et al., 2013) or from Chinese schools (Duan et al., 2010). Study quality was on average 8.30 (*SD* = 1.92; range: 1 to 11) across age groups. It was similar on average for preschool children ($\bar{x}$ = 8.56), school-aged children ($\bar{x}$ = 8.31), adolescents ($\bar{x}$ = 8.00), and adults ($\bar{x}$ = 9.25). It was lower for older adults ($\bar{x}$ = 6.86) due to one study receiving a single study quality point (Frazier et al., 2015). When this outlier was removed, the mean study quality for older adult studies increased to 7.83, which was more similar to the other age bands.

Model fit indices and accepted models. Table 2 provides fit indices for accepted measurement models identified by the systematic review, along with estimated power (based on *N* and *df*; Hancock, 2006), the number of factors, and names of factors included in the accepted model. Considering fit indices, all accepted models had CFI values ≥ .95 and all RMSEA values ≤ .06, indicating excellent statistical fit for the models (Hu & Bentler, 1999). These excellent model fit statistics stood in contrast to the predominantly low power estimates across studies, which came to an average of 0.43 (*SD* = 0.31; range = 0.08 to 0.99). The accepted models included anywhere between one to five factors. Overall, 8 studies accepted a one-factor model (17.78%), 17 accepted a two-factor model (37.78%), 14 accepted a three-factor model (31.11%), 1 accepted a four-factor model (2.22%), 1 accepted a five-factor model (2.22%), and 4 accepted a nested factor model (8.89%). For the calculation of these totals and those reported below, Carlson et al. (2014) was considered to have accepted a one-factor model based on parsimony, although these authors specified no preference between a one-factor or two-factor model; and de Frias et al. (2009) accepted a two-factor model for their Cognitively Normal Subsample, although this model was never formally evaluated.

For preschool samples, roughly half of researchers accepted a one-factor model solution (Number of studies [*k*] = 5; 55.56%; Carlson et al., 2014; Masten et al., 2012; Wiebe et al., 2008, 2012; Willoughby et al., 2012a), while the other half found a two-factor solution (*k* = 4; 44.44%; Lerner & Lonigan, 2014; Miller et al., 2012; Monette et al., 2015; Usai et al., 2014). Among the school-aged samples, the most commonly accepted model was the three-factor model (*k* = 7; 46.67%; Agostino et al., 2010; Arán-Filippetti, 2013, Duan et al., 2010; Lambek & Shevlin, 2011; Lehto et al., 2003; Rose et al., 2012), while a smaller set of studies supported a two-factor (*k* = 4; 26.67%; Brocki & Tillman, 2014; Lee et al., 2012, 2013; van der Ven et al., 2013) or one-factor solution (*k* = 3; 20%; Brydges et al., 2012; Xu et al., 2013). One study involving a school-aged sample supported a model best categorized as a nested factor model (*k* = 1; 6.67%; van der Sluis et al., 2007), although these researchers did not label it as such. Among the three adolescent studies, researchers reported a single nested factor model (*k* = 1; 33.33%; Friedman et al., 2011) and a pair of three-factor models (*k* = 2; 66.66%; Lambek & Shevlin, 2011; Xu et al., 2013). For the adult studies, the support was evenly split between a two-factor model (*k* = 2; 25%; Klauer et al., 2010; Was, 2007), a three-factor model (*k* = 2; 25%; Klauer et al., 2010; Miyake et al., 2000), and a nested factor model (*k* = 2; 25%; Fleming et al., 2016; Ito et al., 2015). One study supported a four-factor model (*k* = 1; 12.5%; Chuderski et al., 2012) and another supported a five-factor model (*k* = 1, 12.5%; Fournier-Vicente et al., 2008). The older adult samples predominantly supported a two-factor model (*k* = 5, 62.5%; Bettcher et al., 2016; de Frias et al., 2009; Frazier et al., 2015; Hedden & Yoon, 2006; Hull et al., 2008), while a smaller, but substantial percentage supported a three-factor model (*k* = 3, 37.5%; Adrover-Roig et al., 2012; de Frias et al., 2009; Vaughan & Giovanello, 2010).

Table 3 provides counts and frequencies of how often a specific construct was represented in an accepted factor model. The most common factors were those included in the original measurement model by Miyake and colleagues (2000), with Updating/Working Memory (*k* = 32; 72.73% of models) being the most frequent, followed by Inhibition (*k* = 20; 52.27%), and then by Shifting (*k* = 20; 45.45%). A small number of studies merged these factors, including Inhibition and Shifting (*k* = 5; 11.36%), Inhibition and Updating/Working Memory (*k* = 1; 2.27%), and Shifting and Updating/Working Memory (*k* = 3; 6.82%). Two studies included factors of strategic retrieval or access to long-term memory (*k* = 2; 4.55%; Adrover-Roig et al., 2012; Fournier-Vicente et al., 2008).

Some differences occurred in terms of the factors represented across age spans. A global Executive Function factor was represented among 25% of models (*k* = 11), but constituted a unidimensional factor among children and a nested bifactor among adolescents and adults. No sample beyond the school-aged years provided a unidimensional model solution, and a global Executive Function factor was not observed among any eligible older adult samples. No preschool sample identified shifting as a separate factor, while all three factors were represented in all groups above 6 years of age.

Tests used as indicators. Tables 4 and 5 list the indicators organized by factors for child/adolescent and adult studies, respectively. The division between child/adolescent and adult samples was set at a mean age of 16 years, where those with a mean age at or below 16 years were considered child/adolescent (*k* = 21) and those with a mean age over 16 years were considered adult (*k* = 17). Few studies had a consistent battery of tests for all indicators evaluated, but a small number of measures were common in the evaluation of specific constructs. The tests below are categorized based on either task or paradigm, and do not necessarily indicate that the studies were using the exact same task or the exact same dependent variable deriving from that task. In some contexts, the exact same task or a highly similar task was used across studies (e.g., Digit Span Backward); however, in other contexts, a similar paradigm was used to guide the design of similar, but distinguishable tasks. For example, the Stroop paradigm among children comes in multiple different varieties of tasks, including a Boy-Girl Stroop, Day-Night Stroop, and Color-Word Stroop; all of which involve different stimuli, but similar task demands and load onto inhibition.

The most frequent indicator of inhibition for child/adolescent studies were tasks using the Stroop paradigm (k = 11), followed by tasks using the Go/No-go paradigm (k = 7). Tasks using a Tower paradigm were the third most common indicator for inhibition among child/adolescent studies (k = 4). The most commonly used indicator for updating/working memory was the Digit Span Backward task (k = 7), followed by the Letter-Number Sequencing task (k = 3) and tasks using the *n*-back paradigm (k = 3). For shifting, tasks with card sorting paradigms were the most commonly used as indicators (k = 6), while tasks using a Trail Making paradigm were the second most commonly used (k = 5) and tasks using a verbal fluency paradigm were the third most commonly used (k = 4).

In terms of adult studies, there was a greater frequency at which specific measures were used as indicators across studies. For inhibition, a substantial portion of the adult studies used tasks involving a Stroop paradigm (k = 15), followed by an Antisaccade task (k = 10), and then a Stop-Signal task (k = 7). For updating/working memory, the most frequently used indicators were tasks using the *n*-back paradigm (k = 8) and the Letter Memory task (k = 8), followed by the Keep Track task (k = 6) and Digit Span Backwards task (k = 5). The measurement of shifting was more variable, but still a substantial portion of researchers used the Number-Letter task (*k* = 10), followed by the Plus-Minus task (*k* = 5) and the Local Global task (*k* = 4).

The data extraction protocol involved the extraction of the task names, and did not focus on the specific dependent variables derived from each of these tasks that were ultimately included in measurement models. A brief post-hoc evaluation explored the variety of scores that different researchers used in their models for the most commonly used paradigms: the Stroop task as an indicator for inhibition. The Stroop task consists of congruent/neutral conditions along with incongruent conditions. In congruent/neutral conditions, participants read color words (e.g., blue, red) written in either black ink or their corresponding ink color, or they named the ink color of a non-verbal stimulus (e.g., a line of asterisks or X's). In the incongruent condition, participants see color words written in incongruent ink colors (e.g., blue written in red ink) and they are asked to read the ink color, inhibiting the automatic response of reading the word. Among children, similar tasks use alternative stimuli, such as the Day-Night Stroop where children are shown a sun or moon and asked to say night or day, respectively.

Among the 11 child/adolescent studies using a Stroop-like task, 7 studies included a Stroop Color-Word paradigm, while the remainder involved Day-Night, Boy-Girl or other Stroop-like task. Within the 7 studies using the color-word approach, 6 different dependent variables were identified, including the difference in time-to-completion between the incongruent and neutral/congruent conditions (Agostino et al., 2010; Brydges et al., 2012), the total number correct in the incongruent condition (Arán-Filippetti, 2013), the difference in the number of correct responses between the incongruent and neutral/congruent conditions (Brocki et al., 2014), the median response latency on incongruent trials (Huizinga et al., 2006), the number of items named per second (van der Sluis et al., 2007), and the reaction time difference between incongruent and neutral/congruent conditions (Xu et al., 2013).

Among the 15 studies using a Stroop paradigm among adult samples, 5 different dependent measures deriving from the same test were identified, including a reaction time difference score between incongruent and neutral/congruent conditions (Fleming et al., 2016; Friedman et al., 2011; Fournier-Vicente et al., 2008; Hull et al., 2008; Ito et al., 2015; Klauer et al., 2010; Miyake et al., 2000; Was, 2007), a ratio of proportion correct in the incongruent condition to proportion correct in the neutral/congruent condition (Chuderski et al., 2012), an interference index (de Frias et al., 2009), the total correct in the incongruent condition statistically controlling for the total correct in the neutral/congruent condition (Bettcher et al., 2016; Frazier et al., 2015; Pettigrew et al., 2014), and the reaction time for correct incongruent trials (Vaughan & Giovanello, 2010).

Bootstrapped Re-Analysis

Bootstrapping validation. The bootstrapping re-analysis of the WAIS-IV correlation matrix for the 16-69 year-old sample (N = 1,800) found that the accepted model for the WAIS-IV converged for 100% of the bootstrapped samples, with 100% of samples meeting lenient fit thresholds (i.e., CFI $\geq$.90, RMSEA $\leq$.08). In terms of the strict fit thresholds, 99.76% of bootstrapped samples had a CFI $\geq$.95, but 0% had an RMSEA $\leq$.05. The mean CFI (95% CI) was 0.96 (0.95, 0.97) and the mean RMSEA was 0.06 (0.059, 0.07). The estimated power for this model (df = 79) was 0.99. Using the WISC-V correlation matrix for the full sample (6-16 year-olds; N = 2,200), the accepted model for the WISC-IV converged for 100% of samples, with 100% of these samples meeting the lenient fit thresholds. The strict fit threshold of CFI ≥.95 was met for 94.04% of bootstrapped sample, while 0% of samples met the strict RMSEA ≤.05 cutoff. For the WISC-V, the mean RMSEA was 0.06 (0.053, 0.06) and the mean CFI was 0.95 (0.948, 0.96). The estimated power for this model (*df* = 92) was also 0.99.

Executive function measurement models. As noted earlier, a total of 21 samples met eligibility criteria for the re-analysis. These samples were not evenly divided between the age bands used to divide the studies in the qualitative synthesis: preschool (*k* = 2), school-age (*k* = 8), adolescent (*k* = 2), adult (*k* = 5), and older adult (*k* = 4). Due to the wide span of ages, the samples were stratified into two samples with 16 years of age as the cut point, where 10 samples were considered adult (i.e., >16 years of age) and 11 samples were considered child and adolescent (i.e., ≤16 years of age). Among the child/adolescent studies, the choice was made to exclude the 2 re-analyzed preschool samples from the calculation of summary statistics for that age range (e.g., mean/median percent convergence, mean/median percent meeting fit criteria). This decision was based on (a) the observation that no separate shifting factor was observed for preschool samples in the qualitative synthesis, (b) the extensive literature detailing the early childhood years as unique and fundamental for executive function development (Müller & Kerns, 2015), and (c) the conceptualization of shifting as an ability that arises later in executive function development (Garon, Bryson, & Smith, 2008). The exclusion of the preschool samples led to an age span of child/adolescent studies ranging from 8.33 to 14.41 years composed of 9 samples. The age span for the adult studies ranged from 17.30 to 72.24. The 17-year-old sample (Friedman et al., 2011) was included with the other adult sample due to factor analytic research observing stability of the structure of executive functions from this age into early adulthood (Friedman et al., 2016). Older adults were included within this age band because (a) there was an insufficient number older adult studies to compose its own group; and (b) although there is evidence for age-related declines in performances on executive function tasks (Reynolds & Horton, 2008), the qualitative findings did not provide definitive evidence for de-differentiation. Unlike the preschool age band, all three constructs were represented among this age group, and the oldest sample evaluated produced a three-factor solution (Vaughan & Giovanello, 2010).

Percent convergence. Tables 6 and 7 list the percentage of models that converged among the 5,000 bootstrapped samples for each measurement model specified for child/adolescent and adult studies, respectively. The percent convergence is presented for each individual study, and a mean and median percent convergence is presented for all studies. These summary statistics for percent convergence are visually presented in Figures 3 and 4 for child/adolescent and adult studies, respectively. For both the child/adolescent and adult studies, the rates of convergence were related to model complexity, where models with more parameters tended to properly converge less often; however, the more complex set of models differed across age spans in terms of their frequency of convergence. For example, among adult studies, there was a clear negative relationship between percent convergence and model complexity. The bifactor model converged the least often ($\bar{x}$ = 24%; *Mdn* = 11%). The nested factor ($\bar{x}$ = 50%; *Mdn* = 34%) and three-factor models ($\bar{x}$ = 46%; *Mdn* = 40%) converged infrequently and less often than the three two-factor models, which all converged at roughly the same rate: inhibition-shifting merged ($\bar{x}$ = 76%; *Mdn* = 86%), inhibition-updating merged ($\bar{x}$ = 69%;

Mdn = 69%), and shifting-updating merged ($\bar{x}$ = 68%; *Mdn* = 66%). The unidimensional model converged for almost every bootstrapped sample ($\bar{x}$ = 99%; *Mdn* = 95%).

In contrast to the adult studies, the frequency of convergence among the child/adolescent samples was slightly different, where the model that converged the least often was the three-factor model ($\bar{x}$ = 36%; *Mdn* = 26%), while the nested factor ($\bar{x}$ = 59%; *Mdn* = 60%) and bifactor models ($\bar{x}$ = 48%; *Mdn* = 49%) converged at closer frequencies. For the three two-factor models, the models merging the shifting factor tended to converge more often. The inhibition-shifting merged ($\bar{x}$ = 76%; *Mdn* = 89%) and shifting-updating merged models ($\bar{x}$ = 71%; *Mdn* = 56%) converged more often than the inhibition-updating merged model ($\bar{x}$ = 59%; *Mdn* = 55%). As with the adult studies, the unidimensional model converged for almost every bootstrapped sample ($\bar{x}$ = 97%; *Mdn* = 100%).

Percent of converged models meeting fit criteria. Tables 6 and 7 list the percentage of the converged models that met lenient and strict fit thresholds for each measurement model specified for child/adolescent and adult studies, respectively. The trend in terms of meeting fit thresholds was generally in the opposite direction of model convergence, where the more complex models tended to fit better than the simpler models. This was true for both the CFI and RMSEA, and the trend is visually represented in Figures 5 and 6 for child/adolescent and adult studies, respectively. As also clearly demonstrated by these figures, the strict fit thresholds were rarely met for most models, whereas the lenient fit thresholds, though met more often, were still met infrequently.

For the adult studies, the bifactor model met lenient (CFI: $\bar{x}$ = 63%; *Mdn* = 55%; RMSEA: $\bar{x}$ = 61%; *Mdn* = 60%) and strict fit criteria (CFI: $\bar{x}$ = 36%; *Mdn* = 30%;

RMSEA: $\bar{x}$ = 24%; *Mdn* = 25%) the most often among the bootstrapped samples for which this model converged. The nested factor model met lenient (CFI: $\bar{x}$ = 52%; *Mdn* = 52%; RMSEA: $\bar{x}$ = 56%; *Mdn* = 58%) and strict fit criteria (CFI: $\bar{x}$ = 19%; *Mdn* = 14%; RMSEA: $\bar{x}$ = 17%; *Mdn* = 14%) at roughly the same rate that the three-factor model met lenient (CFI: $\bar{x}$ = 48%; *Mdn* = 44%; RMSEA: $\bar{x}$ = 57%; *Mdn* = 57%) and strict fit criteria (CFI: $\bar{x}$ = 19%; *Mdn* = 10%; RMSEA: $\bar{x}$ = 17%; *Mdn* = 15%). The two-factor models all met the fit criteria at about the same frequency, although the inhibition-updating merged model met the $\leq$.08 RMSEA criterion ($\bar{x}$ = 36%; *Mdn* = 36%) at a greater rate than the other two-factor models, as made visually evident by a peak in the forest plot line in Figure 6.

For the child/adolescent studies, the bifactor met lenient (CFI: $\bar{x}$ = 64%; *Mdn* = 71%; RMSEA: $\bar{x}$ = 50%; *Mdn* = 52%) and strict fit criteria (CFI: $\bar{x}$ = 39%; *Mdn* = 42%; RMSEA: $\bar{x}$ = 21%; *Mdn* = 21%) the most often among the bootstrapped samples for which this model converged. The three-factor model tended to meet lenient and strict fit criteria at about the same frequency as the nested factor model. Similarly, the two-factor models all tended to meet lenient and strict fit criteria at roughly the same rate, while the unidimensional model met lenient (CFI: $\bar{x}$ = 36%; *Mdn* = 48%; RMSEA: $\bar{x}$ = 32%; *Mdn* = 21%) and strict fit criteria (CFI: $\bar{x}$ = 11%; *Mdn* = 6%; RMSEA: $\bar{x}$ = 11%; *Mdn* = 5%) the least often.

The percent of converged samples meeting fit thresholds cannot be properly understood without appreciating the percent of models converging among the bootstrapped samples. Those models that did not converge did not provide fit indices to contribute to this overall estimate, indicating that the percent of fitting model based on fit thresholds alone may over-estimate how often these models were accepted among the 5,000 bootstrapped samples. In turn, the next section presents how often models both converged and met fit criteria among the 5,000 bootstrapped samples across studies.

Percent of models both converging and meeting fit criteria. Among the 5,000 bootstrapped samples for each study, the frequency at which models both converged and met fit criteria was quite low across different models estimated, although some models tended to be more successful than others. The percent of samples for which a specified model both converged and met fit criteria is provided for multiple fit thresholds in Tables 6 and 7 for children/adolescent and adult samples, respectively. Figures 7 and 8 offer a visual representation of these values. These values constitute the percent of samples in which this model would be accepted by a researcher, in that the model both properly converged and met criteria indicative of good model fit.

Among the adult studies, the rate at which models were deemed acceptable was quite low based on lenient fit criteria and extremely low based on strict fit criteria. The nested factor model was the most often accepted model based on both the lenient (CFI: $\bar{x}$ = 33%; *Mdn* = 14%; RMSEA: $\bar{x}$ = 34%; *Mdn* = 17%) and strict fit indices (CFI: $\bar{x}$ = 13%; *Mdn* = 5%; RMSEA: $\bar{x}$ = 11%; *Mdn* = 6%). Based on lenient fit indices, the three-factor model was the second most often accepted model (CFI: $\bar{x}$ = 25%; *Mdn* = 12%; RMSEA: $\bar{x}$ = 32%; *Mdn* = 19%); however, based on strict fit indices, the bifactor model (CFI: $\bar{x}$ = 11%; *Mdn* = 4%; RMSEA: $\bar{x}$ = 8%; *Mdn* = 3%) was accepted more often than the three-factor model (CFI: $\bar{x}$ = 8%; *Mdn* = 5%; RMSEA: $\bar{x}$ = 7%; *Mdn* = 3%). The two-factor models did not differ from the three-factor model or each other in how often they were accepted based on strict fit criteria; however, based on lenient fit criteria, the inhibition-updating merged model was the most often accepted of the two-factor models (CFI: $\bar{x}$ = 16%; *Mdn* = 6%; RMSEA: $\bar{x}$ =30%; *Mdn* = 15%), but it was still accepted less often than the three-factor model. The unidimensional model was comparable to the two-factor models in terms of strict fit criteria, and was very rarely accepted based on lenient fit criteria as well (CFI: $\bar{x}$ = 8%; *Mdn* = 0%; RMSEA: $\bar{x}$ = 13%; *Mdn* = 2%).

The child/adolescent studies did not follow the same trend as the adult studies. As clearly presented in Figure 7, no model stood out as the most often accepted. Instead the inverse occurred, where two models were more frequently *not* accepted, specifically – based on lenient fit criteria – the inhibition-updating merged model (CFI: $\bar{x}$ = 20%; *Mdn* = 20%; RMSEA: $\bar{x}$ = 13%; *Mdn* = 12%) and the three-factor model (CFI: $\bar{x}$ = 21%; *Mdn* = 10%; RMSEA: $\bar{x}$ = 11%; *Mdn* = 8%) rarely converged and met fit thresholds. Based on lenient fit criteria, there was no clear delineation between the unidimensional (CFI: $\bar{x}$ = 36%; *Mdn* = 48%; RMSEA: $\bar{x}$ = 32%; *Mdn* = 21%), shifting-updating merged (CFI: $\bar{x}$ = 35%; *Mdn* = 31%; RMSEA: $\bar{x}$ = 25%; *Mdn* = 32%), inhibition-shifting merged (CFI: $\bar{x}$ = 34%; *Mdn* = 32%; RMSEA: $\bar{x}$ = 27%; *Mdn* = 30%), nested factor (CFI: $\bar{x}$ = 31%; *Mdn* = 26%; RMSEA: $\bar{x}$ = 21%; *Mdn* = 18%), or bifactor models (CFI: $\bar{x}$ = 28%; *Mdn* = 23%; RMSEA: $\bar{x}$ = 20%; *Mdn* = 22%). There was a bit more of a distinction based on strict fit criteria, where the nested factor (CFI: $\bar{x}$ = 17%; *Mdn* = 13%; RMSEA: $\bar{x}$ = 7%; *Mdn* = 4%) and bifactor models (CFI: $\bar{x}$ = 16%; *Mdn* = 13%; RMSEA: $\bar{x}$ = 8%; *Mdn* = 9%) were more often accepted based on CFI, but this trend was not evident based on the RMSEA, which takes model complexity into account.

***Mean fit indices and inter-factor correlations*.** Tables 8 and 9 provide the mean fit indices (i.e., CFI and RMSEA) and 95% confidence intervals for child/adolescent and adult studies, respectively. These statistics are based only on the models that converged and provided an estimate of the fit indices. For all models that converged involving correlated factors, Tables 10 and 11 for child/adolescent and adult studies, respectively, provide the mean inter-factor correlations and 95% confidence intervals.

Post-hoc evaluation of publication bias. The re-analysis focused on model convergence and fit regardless of which model was originally accepted for each individual study. A post-hoc analysis evaluated the presence of publication bias by examining the rate of model acceptance among the 5,000 bootstrapped samples for the model originally accepted by the researchers using their observed sample. This analysis was done using only those studies with accepted models that corresponded to those seven evaluated in the re-analysis, which resulted in 10 child/adolescent samples and 8 adult samples. Although these values are present in Tables 6 and 7, they are presented in isolation in Table 12 as well for the convenience of the reader. Among child/adolescent studies, the rate at which the originally accepted models would be selected among the 5,000 bootstrapped sample was low using both lenient fit criteria (CFI: $\bar{x}$ = 36%; *Mdn* = 43%; RMSEA: $\bar{x}$ = 33%; *Mdn* = 31%) and strict fit criteria (CFI: $\bar{x}$ = 15%; *Mdn* = 15%; RMSEA: $\bar{x}$ = 13%; *Mdn* = 12%). Among adult studies, this rate was also low using lenient (CFI: $\bar{x}$ = 27%; *Mdn* = 12%; RMSEA: $\bar{x}$ = 34%; *Mdn* = 16%) and strict fit criteria (CFI: $\bar{x}$ = 6%; *Mdn* = 3%; RMSEA: $\bar{x}$ = 7%; *Mdn* = 4%).

Discussion

The systematic review and re-analysis summarized an extensive body of research exploring executive functions over the last two decades, identifying a large set of studies producing occasionally inconsistent findings about the structure of executive functions over the course of the lifespan. A qualitative synthesis of this research covered sample demographics, test selection, study quality, model fit, and the frequency at which different constructs and models appeared in the published literature. The majority of samples identified were composed of children and adolescent s (k = 27), while a smaller portion of studies involved adults (k = 8) and older adults (k = 8). In terms of the factor models supported by eligible studies, there was evidence for increasing multidimensionality of executive functions over the course of development. Preschool samples were roughly split between a one-factor and two-factor solution, with no studies identifying a specific shifting factor. School-aged samples showed more support for a three-factor model than a two-factor model, while the adolescent samples supported three-factor and nested factor solutions. There was equal support for two-factor, three-factor, and nested factor models among adult samples. One of the studies producing a two-factor solution among adults did not test a three-factor solution (Was, 2007), and the other involved two studies and found a three-factor solution in their second study (Klauer et al., 2010). Combined these findings indicate a gradual differentiation of executive functions from preschool into adulthood, and the potential emergence of a specific shifting factor around school-age to adolescence.

In terms of the consistency between adult and older adult studies, most older adult studies supported a two-factor solution, but there was also support for a three-factor solution. These findings could indicate a slight de-differentiation of abilities with older age; however, no studies supported a one-factor solution, a three-factor solution was supported in the oldest sample evaluated (Vaughan & Giovanello, 2010), and – unlike the preschool age group – all three factors were represented in at least one of the measurement models evaluated within this age band. As well, researchers have yet to evaluate the structure of executive function for a substantial portion of mid-life: none of the samples evaluated had a mean age between 30 and 60 years. In turn, if executive functions do de-differentiate, the representation of ages within the current review is not comprehensive enough to identify the time of life at which this de-differentiation occurs, indicating the need for more research on samples in middle adulthood along with more longitudinal investigations. The only longitudinal study evaluating changes in executive functions among older adults involved just two time points separated by a three-year interval among adults already aged 55 years and above (de Frias et al., 2009). Overall, the results from this systematic review and re-analysis do not support the de-differentiation of executive function with older age, due largely to insufficient longitudinal evidence and large gaps in the age spans represented in cross-sectional research.

Based the qualitative synthesis in this systematic review, there was no consensus between studies within any of the age ranges in regards to the structure of executive functions. Most of the studies included in the qualitative synthesis were of good quality (e.g., 80% of studies had a study quality score of $\geq 8/11$), although very few had sufficient power (e.g., 20% $\hat{\pi} \geq .80$). Despite low power, all published studies reported excellent fit for their models (i.e., CFI $\geq .95$; RMSEA $\leq .06$), which provides no means for a reviewer of the overall literature to preferentially select one model from one study over an equally well-fitting model from another study.

The excellent fit and low power of the published models brings into question whether those models that fit well among a specific sample and specific battery of tests happen to be the models that get published, while other models that do not meet standard fit cutoffs remain in the file drawer. This concern aligns with the general concern of replicability currently facing psychological science (e.g., Pashler & Harris, 2012; Pashler & Wagenmakers, 2012; Simmons, Nelson, & Simonsoh, 2011). A good fitting model captures the data well, but it does not necessarily reflect the true model for the population (Hancock, 2006). Considering the low power of these excellent fitting models, the question remains whether they could be replicated among small samples drawn from the same population. The majority of studies were underpowered and denoted as conceptual replications, rather than direct replications using identical test batteries and recruiting a sufficient sample size. These studies often found similar results to the first measurement model of executive functions (Miyake et al., 2000) despite using a different collection of tests and often an alternative population from which to sample. As with direct replication failures, conceptual replication failures are rarely published (Makel, Plucker, & Hegarty, 2012). In turn, it is possible that the many published studies that contain the most frequently reported factors (i.e., inhibition, updating and shifting) may be the conceptual replication successes, while the failures not supporting a three-factor model remain in the file drawer.

A post-hoc analysis shed light on the issue of publication bias and potential non-replicability within this field. On average, the accepted models reported by researchers were selected among only about a third of bootstrapped child/adolescent (i.e., 33-36%) and adult samples (i.e., 27-34%) based on lenient fit criteria. These findings clearly illustrate a substantial publication bias across studies reporting measurement models for executive functions. This bias affected the results of the re-analysis, which found low rates of model acceptance for all the models evaluated, although some models appeared to fit the data more consistently than others.

The re-analysis effort aimed to explore how well seven alternative models fit the data across multiple samples and test batteries. As noted, the most telling findings from this re-analysis was the remarkably low rate at which many published models converged and/or met fit thresholds among bootstrapped samples. This method was validated by an evaluation of the WAIS-IV and WISC-V models, which converged and met fit criteria for nearly every bootstrapped sample. For the executive function test batteries, the most frequently accepted factor model among adults was the nested factor model. This model includes two specific factors for updating and shifting, but no specific factor for inhibition, along with a common executive function bifactor on which all indicators load or co-load. In comparison to the WAIS-IV and WMS-IV models, the nested factor model only converged 50% of the time on average across samples. Among those samples for which the model converged, only 56% had an RMSEA $\leq$ 0.08 and only 52% had a CFI $\geq$ 0.90. In turn, despite being the most often accepted, the nested factor model would be accepted among only 33 to 34% of 5,000 bootstrapped samples on average across studies.

The low rate of model convergence may derive in part from low construct reliability of factors included in the model, where a limited amount of true construct variance is present for the factors specified (Gagne & Hancock, 2006). As observed in previous re-analyses of executive function measurement models, factors within this field often have weak to moderate levels of reliability (Willoughby et al., 2014), indicating limited construct-specific variance captured by latent factors. In the current re-analysis, the models that fit the least often on average were those with the most factors. For example, the bifactor fit very rarely, judging there needed to be sufficient unique variance in the common factor and all specific factors to ensure adequate construct reliability and non-zero loadings. In the original selection of a nested factor model, the decision to drop the inhibition-specific factor was guided by low loadings onto this factor in the context of a bifactor model (Friedman et al., 2008). Considering the low reliability (Baggetta & Alexander, 2016; Schmidt, 2003) and low inter-test correlations often observed for executive function tests (Miyake et al., 2000), the manifest variables included in the re-analysis could have had limited construct variance related to the factor(s) on which they loaded (Müller & Kerns, 2015). In turn, during the re-analysis effort, there may be insufficient construct-specific variance in the data for many of the models to properly converge.

A clear relationship existed between model complexity and convergence, in that more complex models converged less often. This same relationship existed for model fit, where more complex models better fit the data. When interpreting the re-analysis findings, these conflicting patterns made model selection a difficult task. While a unidimensional model almost always converges, it will almost never adequately fit the data among adults. In contrast, a nested factor model rarely converges, but when it does, it will more often meet traditional fit thresholds.

Despite issues of publication bias and low model convergence, the published results offer some empirical information about the nature of executive functions. The statistician George Box, once wrote “all models are wrong, but some are useful,” (Box & Draper, 1987, p. 424), which applies well to the current findings. For the adult studies, three of the highest quality studies accepted the nested factor model using the same test battery across different samples (Fleming et al., 2016; Friedman et al., 2011; Ito et al., 2015). The results of these three studies align with the results of the overall re-analysis; however, there was still variability around how often this model was accepted across these studies based on the re-analysis. The convergence rate ranged from 22% to 96% and the acceptance rate ranged from 16% to 84% and 14% to 86% for the lenient cutoffs of the CFI and RMSEA, respectively. Even within a small set of consistent studies with well-powered, similarly aged samples (all $\hat{\pi}$ range: 0.74 to 0.99; $\bar{x}$ age range: 17.30 to 22.50), there was pronounced variability in how often the nested factor model was accepted among the 5,000 bootstrapped samples.

Although the cumulative evidence points towards the nested factor model, it is not well differentiated from the three-factor model using lenient fit thresholds, which was accepted among the bootstrapped samples at a comparable rate based on the RMSEA (i.e., nested factor: $\bar{x}$ = 34%, *Mdn* = 17%; three-factor: $\bar{x}$ = 32%, *Mdn* = 19%), but not as much the CFI (i.e., nested factor: $\bar{x}$ = 33%, *Mdn* = 14%; three-factor: $\bar{x}$ = 25%, *Mdn* = 12%). Because the RMSEA favors parsimony (Hooper et al., 2008), the comparable rates of acceptance between the nested factor model and three-factor model based on this fit index could indicate that the nested factor model is too complex, with limited improvement in fit despite increased model complexity. The same argument could be made, however, for the inhibition-updating merged model, which had a similar rate of acceptance based on the RMSEA (i.e., $\bar{x}$ = 30%, *Mdn* = 15%), but a much lower rate based on the CFI (i.e., $\bar{x}$ = 16%, *Mdn* = 6%). However, the only previous publication to support such a model (Klauer et al., 2010) had just two indicators on their inhibition and updating factors, and reported a second study within the same publication that found a three-factor solution using a more extensive battery of tests. In turn, the factor solutions most supported for adult samples by the re-analysis effort were the three-factor and nested factor solutions, without a clear determination about which model should be preferred based on model fit.

A method for determining which measurement model ultimately aligns with the true nature of executive functions will require a closer examination of the brain-behavior relationships that underlie the constructs included in the accepted measurement model. Researchers have found brain activity during performance-based tasks of executive functions in areas associated with specific constructs, including right inferior frontal cortex, basal ganglia, and pre-supplementary motor area activity during inhibition tasks (Aron, 2008), dorsolateral prefrontal cortex activity (DLPFC; Stuss & Levine, 2002) as well as frontopolar activity (Collette et al., 2005) during updating/working memory tasks, and DLPFC and dorsal anterior cingulate cortex activity during shifting tasks (Luna, Marek, Larsen, Tervo-Clemmens, & Chahal, 2015). Although specific brain-behavior relationships have been proposed, there is evidence for both the unity and diversity of brain activity underlying separate executive function constructs (Collette et al., 2005, 2006). A comprehensive meta-analytic investigation (Niendam et al., 2012) found strong evidence for a superordinate fronto-cingulo-parietal network that showed common activity during tasks tapping into inhibition, working memory, and flexibility (i.e., a term often used synonymously with shifting; Baggetta & Alexander, 2016). This integrative function could parallel the common bifactor present in the nested factor model, which past researchers have conceptualized as the ability to "actively maintain task goals and goal-related information and use this information to effectively bias lower-level processing" (Miyake & Friedman, 2012, p. 11), arguably necessary for successful performance across executive function domains.

The two competing models that emerged from the re-analysis of adult samples show some alignment with research on brain-behavior relationships; however, the results of the re-analysis of the child/adolescent samples were far more ambiguous, and interpretable in the opposite fashion to that of the adult studies. Whereas for the adult studies in Figure 8, there was a clear peak in model acceptance rates for the three-factor and nested factors models, the child/adolescent studies in Figure 7 had two definitive valleys for the inhibition-updating merged and three-factor models, evidencing that models with differentiated shifting factors were *less* preferable to models that either merged the shifting factor or had a strong common executive function bifactor. This trend was consistent with discussion of a non-differentiated shifting factor early in development (Garon et al., 2008) and the notion that an independent shifting ability emerges later in development (Müller & Kerns, 2015). This trend was observed despite removing preschool samples from the measures of central tendency calculated in the re-analysis.

The competing child/adolescent models that both converged and exceeded lenient fit thresholds most often were the unidimensional, shifting-updating merged, inhibition-shifting merged, nested factor, and bifactor models. While these models were not easily differentiated based on the lenient CFI cutoff, the lenient RMSEA cutoff was met most often for the unidimensional ($\bar{x}$ = 32%; *Mdn* = 21%), the shifting-updating ($\bar{x}$ = 25%; *Mdn* = 32%), and inhibition-shifting models ($\bar{x}$ = 27%; *Mdn* = 30%). Considering the greater complexity of the nested factor and bifactor models, the more parsimonious models were favored by the RMSEA cutoff. As with the adult studies, there was not a clear determination about which model should be preferred based on fit indices; however, the re-analysis of child/adolescent samples supported (a) either a unidimensional or two-factor solution and (b) a model that does not have a differentiated shifting factor. This non-differentiated system is supported by neurodevelopmental trajectories, where grey matter in the DLPFC, which is associated with both updating/working memory and shifting (Luna et al., 2015; Stuss & Levine, 2002), is pruned after the ventral frontal regions associated with inhibition (Aron, 2008) during child and adolescent development (Müller & Kerns, 2015).

This systematic review and re-analysis offers the first comprehensive and empirical summary of measurement models for executive function test batteries across the lifespan. Despite the comprehensiveness of this review, the conclusions drawn from it remain tentative due to a variety of limitations. A first limitation pertains to the limited diversity of the samples evaluated. The eligible samples were largely balanced in gender (i.e., 52.56% female); however, the samples were not diverse in terms of their ethnic and racial composition. Ethnic or racial demographics were only reported for 40% of samples, with clear discrepancies across age ranges in terms of how often this information was reported. Although 66% of preschool samples had racial or ethnic makeup reported, only 25% of older adult studies provided similar information. There were some studies with specifically Chinese samples (Duan et al., 2010; Lee et al., 2012; Xu et al., 2013) or majority minority samples (Masten et al., 2012; Rose et al., 2012); however, these few ethnically and racially diverse samples were exclusively child and adolescent.

Based on reported demographics, the adult and older adult samples were not only mostly White, but were also highly educated. Half of the adult samples were undergraduate populations, while the older adults ranged in education from 11.30 to 17.67 years, with all but one sample having over 15 years of education on average. Based on the sample demographics, the generalizability of this research to diverse populations remains questionable. Furthermore, although the mean ages ranged from 3.01 to 73.68 across samples, there was still a gap in the representation of middle adulthood. As noted earlier, no researchers reported a sample with a mean age between 30 and 60. In turn, the structure of executive functions within middle adulthood remains largely unevaluated, because most studies categorized as adults in this review evaluated an undergraduate or college-aged sample. Future researchers would benefit from recruiting more participants within middle adulthood, without post-secondary education, and from diverse ethnic or racial backgrounds. This would ensure that the research findings on the structure of executive functions are representative beyond a well-educated and White population.

Additional limitations pertained specifically to the re-analysis efforts, which made some analytical decisions and assumptions that ultimately limit the interpretation of the current findings. Specifically, the wide age ranges used for both the child/adolescent ($\bar{x}$ age range: 8.33 to 14.41) and adult samples ($\bar{x}$ age range: 17.30 to 72.24) limit inference about the structure of executive functions at specific points in human development (e.g., childhood vs. adolescence, young vs. older adulthood). Collapsing across developmental periods ensured a roughly equal number of samples fell within the child/adolescent (k = 9) and adult (k = 10) age spans prior to calculating a mean and median for rates of convergence and model acceptance. Developmental considerations were taken prior to calculating measures of central tendency, such as excluding preschool samples due to a non-differentiated shifting factor (Miller et al., 2012; Usai et al., 2014). Despite wide age bands, conclusions based on a larger collection of samples arguably allow for more accurate inference about the structure of executive functions during development and adulthood.

Another limitation of the systematic review was the lack of individual participant data, because the findings presented in the re-analysis were based solely on simulated data using correlation matrices. Non-parametric bootstrapping with re-sampling is a more common method used by researchers with their raw datasets, but was not possible using summary data. If researchers were to use non-parametric bootstrapping with re-sampling to re-analyze their own sample data, the conclusions may differ from those amalgamated in the current review. In the context of the re-analysis, the parametric bootstrapping simulates samples of the same *N* as the observed samples, pulled from an assumed multivariate normal distribution. The alternative non-parametric bootstrapping with re-sampling approach more commonly used with raw data would not make this assumption; and software packages commonly used for confirmatory factor analysis would not offer a confidence interval around fit indices, nor a rate at which simulated samples met fit cutoffs. However, some software packages (e.g., MPlus; Muthén & Muthén, 2014) would quantify the number of bootstrapped draws completed, which would give an estimate of how often the model would properly converge. The use of bootstrapping may be fruitful for future researchers to guide their model selection, allowing them to determine the frequency at which an excellent fitting model would replicate among a set of bootstrapped samples.

In terms of future directions for researchers evaluating measurement models of executive functions, many gaps in the field remain unresolved based on the current review. As is clear from the findings, the results provided some guidance in regards to which models have the most – or least – empirical support, but they did not suggest that any model should be unequivocally accepted. Future researchers should evaluate alternative models including factors not previously represented in published measurement models. Despite some inconsistencies in the naming of factors, most researchers have taken the approach of evaluating the three-factor model (i.e., inhibition, updating, and shifting; Miyake et al., 2000), which has substantially influenced their test selection and design. Just a small set of studies explored additional constructs (e.g., Access to Long Term Memory, Adrover-Roig et al. (2012); Hot and Cool Executive Function, Carlson et al., 2014; Strategic Retrieval, Fournier-Vicente et al., 2008). Future researchers should consider exploring new constructs that have been postulated in previous research, but not consistently evaluated in confirmatory factor analyses, such as planning, problem solving, fluency, and reasoning (Packwood et al., 2011). As well, factor analytic studies not covered in this review have explored the multidimensionality of specific executive function constructs (e.g., inhibition, Aichert et al., 2012, Friedman & Miyake, 2004; problem solving, Cinan et al., 2013; Scherer & Tiemann, 2014), indicating that sub-components under the umbrella term of executive functions may be umbrella terms within themselves and worth further exploration.

In addition to the measurement of different constructs, other methods for advancing the field could include either the re-analysis of primary datasets or the addition of longitudinal follow-ups to research designs. One recent re-analysis explored a formative factor model as an alternative method of both modeling and interpreting performances on tests of executive functions (Willoughby & Blair, 2016). While a formative model simply flips the directional path between manifest variables and factors (Kline, 2006), other re-analyses could conceptualize executive functions in a more causal manner. If conceptualizations of executive functions in early childhood suggest that inhibition and updating precede shifting development (Garon, 2008), then an alternative model could use causal paths, where shifting is endogenous to inhibition and updating in a structural equation modeling framework. In terms of longitudinal follow-up, only a small set of studies have evaluated longitudinal invariance of executive function factors (e.g., de Frias et al., 2009; Friedman et al., 2016; Lee et al., 2013; Willoughby et al., 2012b), and future longitudinal research designs may clarify which factor structures are stable and replicable over time.

Another important consideration in future research designs is the consistency of the tests used to measure executive functions. While some tests were used consistently (e.g., the Stroop task, Antisaccade, *n*-back), a post-hoc exploration of the Stroop paradigms identified inconsistencies in the dependent variables that were derived from Stroop tests and ultimately used as indicators in measurement models. There were six different dependent variables deriving from Stroop paradigms among child/adolescent studies, and five different dependent variables deriving from Stroop paradigms among adult studies. Aside from variability in the exact scores used across confirmatory factor analyses, there was substantial variability in the batteries used across studies as well. Although there is some consistency in the indicators assigned to different constructs, few studies had the exact same test battery, which could explain the inconsistencies in factor solutions and inter-factor correlations across different studies. Three of the highest quality studies were based on a common test battery (Fleming et al., 2016; Friedman et al., 2011; Ito et al., 2015), and all three accepted the nested factor model. The factor structure of this battery has also been evaluated longitudinally, showing stability in its structure over a 6-year period (Friedman et al., 2016).

While few studies in the executive function literature have used a consistent test battery, the evaluation of executive functions in clinical practice is similarly disparate (Rabin et al., 2016). Since the first published measurement model on executive function, there has been a push for the translation of latent variable research into clinical practice (Miyake, Emerson, & Friedman, 2000), but practitioners do not often use composite scores of executive functions in their assessments. The continued evaluation of executive functions in both academic and clinical settings will require consistent measurement in order to provide comparable and interpretable results, which is a primary focus of the second chapter in this dissertation; however, any consensus in regards to its measurement would likely require an updated review of the many tests used to measure specific constructs to date (Chan et al., 2008), and a gathering of top researchers in the field to arrive at a preferred battery with a strong psychometric foundation to rationalize its widespread use (Baggetta & Alexander, 2016). Some researchers have attempted to produce batteries for widespread dissemination. The National Institute of Health funded the development of a test battery for the assessment of executive functions in clinical trials (i.e., Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research, EXAMINER; Kramer et al., 2014), providing factor scores for working memory, fluency, cognitive control, and a global composite, which align at least partly with the factors supported by the re-analysis of adult samples. Researchers have begun to closely focus on the measurement of executive functions; however, the field is still largely inconsistent in the exact methods recommended to measure the many constructs falling under this umbrella term.

Although this systematic review and re-analysis does not elucidate one preferred model of executive functions, it does show support for a one or two factor solution among children and young adolescents and a three or nested factor solution among adults. While these findings could guide test development, they are particularly tentative and do not offer any definitive conclusions regarding the true nature of executive functions. Alternatively, the findings provided herein offer an affirmation of the “elusive nature of executive functions” (Jurado & Rosselli, 2007, p. 213). Despite decades of measurement, cognitive and neuropsychological researchers have yet to define and measure this construct in a consistent and evidence-based manner. This is not for lack of effort, as made clear by the many studies reviewed herein; and the many more studies evaluating executive functions, but beyond the scope of this review. Nonetheless, despite its elusive nature, the goals of defining, measuring, and understanding executive functions remain tantamount to psychological research, considering the many clinical and societal outcomes associated with executive functions (e.g., Espy et al., 2011; Karr, Areshenkoff, & Garcia-Barrera, 2014; Moffitt et al., 2011; Snyder, 2013; Scott et al., 2015) and the interventions already developed to improve executive functions across the lifespan (e.g., Baggetta & Alexander, 2016; Diamond & Lee, 2011; Karr, Areshenkoff, Rast, & Garcia-Barrera, 2014; Krasny-Pacini, Chevignard, & Evans, 2014).

Chapter 2: Examining the Latent Structure of the Delis-Kaplan Executive Function System

Abstract

Using the Delis-Kaplan Executive Function System (D-KEFS), the current study aimed to (a) determine whether the D-KEFS taps into three executive function factors (i.e., inhibition, shifting, and fluency) and (b) assess the relationship between these factors and tests of executive-related constructs less often measured in latent variable research (e.g., reasoning, abstraction, and problem solving). Participants included 425 adults from the D-KEFS standardization sample (20-49 years-old; 50.1% female; 70.1% White). Eight alternative measurement models were compared based on model fit, with test scores assigned *a priori* to three factors: inhibition (i.e., Color-Word Interference, Tower), shifting (i.e., Trail Making, Sorting, Design Fluency), and fluency (i.e., Verbal/Design Fluency). The Twenty Questions, Word Context, and Proverb Tests were predicted in separate structural models. The three-factor model fit the data well (CFI=0.938; RMSEA=0.047), although a two-factor model, with shifting and fluency merged, fit similarly well (CFI=0.929; RMSEA=0.048). A bifactor model fit best (CFI=0.977; RMSEA=0.032) and explained the most variance in shifting indicators, but rarely converged among 5,000 bootstrapped samples. When the three first-order factors simultaneously predicted the criterion variables, only shifting was uniquely predictive (p<0.05; R^2=0.246-0.408). The bifactor significantly predicted all three criterion variables (p<0.001; R^2=0.141-242). The findings support a three-factor D-KEFS model (i.e., inhibition, shifting, and fluency), although shifting and fluency were highly related (r=0.696). The bifactor showed superior fit, but converged less often than other models. Shifting best predicted tests of reasoning, abstraction, and problem solving. These findings support the validity of D-KEFS scores for measuring executive-related

constructs and provide a framework through which clinicians can interpret D-KEFS results.

Introduction

The debate surrounding the unity and diversity of frontal lobe functions began decades ago (Teuber, 1972), with this debate more recently applied to the nature of executive functions (Miyake et al., 2000; Miyake & Friedman, 2012), an umbrella term for higher-order cognitive abilities related to the functioning of the frontal cortex in addition to connected brain areas (e.g., anterior cingulate, parietal cortex, basal ganglia, cerebellum; Niendam et al., 2012). Although some early conceptualizations of executive or supervisory cognitive systems appeared unitary (e.g., Baddeley & Hitch, 1974; Norman & Shallice, 1986), the first definition for the term executive functions described a multidimensional system, involving "four components" (e.g., goal formulation, planning, carrying out goal-directed plans, and effective performance; Lezak, 1983, p. 507), while sequential definitions have continued to adopt multidimensional conceptualizations of executive functions (Baggetta & Alexander, 2016; Barkley et al., 2012). Consequentially, the terms used to describe diverse executive functions (Packwood, Hodgetts, & Tremblay, 2011) and the tests purported to examine executive functions (Chan, Shum, Toulopoulou, & Chen, 2008) have gradually increased since the construct was first defined. The resulting body of literature presents an array of executive functions, with multiple factors representing dissociable, but related abilities involved in the self-regulation of cognition and behavior (Jurado & Rosselli, 2007).

To quantify the multidimensional nature of executive functions, researchers have applied multivariate statistical methods to identify the latent dimensions of test batteries examining executive functions (Royall et al., 2002). At the turn of the millennium, a seminal article in this line of research reported a confirmatory factor analysis that supported three latent variables based on the common construct variance in their test battery (i.e., updating, shifting, and inhibition; Miyake et al., 2000). Since the publication of this article, the multivariate literature on executive functions has burgeoned (as extensively reviewed in Chapter 1), with numerous sequential latent variable studies examining executive functions during childhood (e.g., Agostino, Johnson, & Pascual-Leone, 2010; Huizinga, Dolan, & Van der Molen, 2006; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003), adolescence (e.g., Friedman, Corley, Hewitt, & Wright, 2009), young adulthood to mid-life (e.g., Chuderski, Taraday, Necka, & Smolen, 2012; Fournier-Vicente, Larigauderie, & Gaonac'h, 2008; Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010) and later age (e.g., de Frias, Dixon, & Strauss, 2006, 2009; Vaughan & Giovanello, 2010).

The advantage of confirmatory factor analysis in the assessment of executive functions derives from the inherent impurity of tests purported to measure executive functions (Miyake & Friedman, 2012). As control processes, executive functions operate through lower-level abilities, meaning that non-executive cognitive functions contribute to performances on executive-related neuropsychological measures (Burgess, 1997; Duggan & Garcia-Barrera, 2015; Hughes & Graham, 2002; Phillips, 1997). A confirmatory factor analytic approach results in latent variables based on shared variance between tests of executive functions, providing a purer estimate of these higher-order cognitive abilities than any individual neuropsychological test (Miyake et al., 2000).

Despite the advantage of latent variables in quantifying executive functions, clinicians do not commonly use composite scores based on evidenced-based factors in clinical practice. The three most heavily researched factors of executive function – the updating of working memory, the shifting of mental sets, and the inhibition of prepotent responses (see Chapter 1) – have garnered a significant amount of empirical support through conceptual replications of the first measurement model (Miyake et al., 2000) across different age groups (e.g., Agostino et al., 2010; de Frias et al., 2009; Klauer et al., 2010; Lehto et al., 2003; Friedman et al., 2009; Vaughan & Giovanello, 2010), and more recent research initiatives have also supported inhibition as a more unitary dimension of the executive system (Miyake & Friedman, 2012). However, despite the evidence for these factors, they have only been evaluated in an experimental context, with no published applications of these factors through clinical test batteries.

Based on factor analyses, some composite measures have become available in clinical settings through commercial measures of executive functions (e.g., the Neuropsychological Assessment Battery Executive Functions Module, White & Stern, 2003; the Test of Executive Control, Isquith, Roth, & Gioia, 2009); but these composites do not represent the factors supported by the published literature. More closely aligning with published research, a recent initiative by the National Institute of Health involved the development of a battery of executive function tests for inclusion in clinical trials (i.e., Executive Abilities: Measures and Instruments for Neurobehavioral Evaluation and Research, EXAMINER; Kramer et al., 2014), providing factor scores of diverse executive functions (e.g., working memory, fluency, and cognitive control) and one global composite of unitary executive function. The investment in the EXAMINER evidences the need and demand for evidence-based executive function outcomes in clinical research; a need and demand that transcends research into clinical practice.

Although executive-related composite scores remain uncommon in clinical settings (Rabin, Paolillo, & Barr, 2016), clinicians have moved towards broadband assessments of executive functions by developing batteries that evaluate diverse higher-order abilities through multiple tests (Jurado & Rosselli, 2007). Increasingly common in clinical practice, the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001) stands as a nine-test battery composed of traditional and newly developed measures of executive functions. Past researchers have evaluated its latent structure, identifying some evidence of diverse latent abilities explaining test performances (Latzman & Markon, 2010; Floyd, Bergeron, Hamilton, & Parra, 2010); however, these researchers used largely exploratory approaches, although confirmatory approaches, specifying *a priori* measurement models that consider task impurity and previous empirical findings, have since become the standard of the field when examining the latent structure of executive functions.

Analyzing the correlation matrices published in the D-KEFS technical manual (Delis et al., 2001), Latzman and Markon (2010) used an exploratory factor analysis to identify a three-factor solution for the D-KEFS Total Achievement scores (for all tests except Proverbs) across three age spans (i.e., 8-19, 20-49, and 50-89). These researchers found performances on D-KEFS tests grouping to form test-specific factors rather than construct-specific factors. Only Card Sorting outcomes loaded on the first factor, only Verbal Fluency outcomes loaded on the second factor, and two Color-Word Interference Test (CWIT) outcomes, plus one Trail Making Test (TMT) outcome, loaded on the third factor (i.e., all timed tests). The researchers labeled these factors cognitive flexibility, monitoring, and inhibition, respectively.

A second research team administered 25 tests from the D-KEFS and the third edition of the Woodcock-Johnson Test of Cognitive Abilities (WJ-III) to a sample of 100 children, evaluating whether the latent structure of a combined test battery aligned with the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (Floyd et al., 2010). These researchers found the D-KEFS tasks dispersed across different CHC factors during the exploratory factor analysis, while also identifying support for a second-order general factor. Taking these findings, the researchers ran a confirmatory model of their exploratory factor structure and identified acceptable fit for their model. However, many D-KEFS tests loaded on factors conceptually disparate from the construct that the D-KEFS tests were purported to measure. The D-KEFS CWIT, a version of the Stroop test (i.e., a measure used as an indicator for inhibition in past measurement models; Klauer et al., 2010; Miyake et al., 2000), loaded on the Processing Speed (*Gs*) factor, likely due to a primary outcome of time-to-completion. Three other D-KEFS tests (i.e., Free Sorting, Word Context, and Twenty Questions) loaded on a Comprehension-Knowledge (*Gc*) factor. These three tests tap into higher-order cognition, but likely share variance with *Gc* tasks because of their reliance on crystallized knowledge. When these tests co-loaded on a broad Executive Function factor, the overall model fit improved, which indicates unexplained executive function variance remaining in the model; however, these additional paths were rejected due to a lack of parsimony.

Both past factor analyses on D-KEFS data ignore the task impurity problem endemic in the measurement of executive functions (Burgess, 1997; Hughes & Graham, 2002; Phillips, 1997) and identify factors based largely on common method variance, rather than the variance of executive function constructs. Floyd and colleagues (2010) concluded that “if there are measures of abilities associated with executive functions, they are contaminated by the general factor and more specific ability factors, so that there is probably little unique about them” (p. 734). Notably, in a second-order confirmatory factor model, the relationship between the second-order factor and the manifest variable is fully mediated by the first-order factors (Reise, Moore, & Haviland, 2010). In turn, the executive function variance not accounted for by first-order factors, composed largely of method variance, remains unexplained error variance in the model.

The primary issue with the approach of these researchers is their limited consideration for past theory and research on executive functions. A latent variable analysis of the D-KEFS guided by past empirical evidence would provide an outcome more useful to clinicians using the broadband measure in clinical settings. Although flawed, these past factor analyses evidence (a) the task impurity of the D-KEFS measures – as is common to all tests of executive functions (Burgess, 1997; Hughes & Graham, 2002; Phillips, 1997) – and (b) the latent interrelatedness between these measures. In turn, an interpretation of D-KEFS tests independently from one another provides a biased estimate of executive functions; however, an interpretation that understands performance patterns in aggregate will provide an evaluation of executive functions more in alignment with current research on the construct, so long as the interpretation considers the influence of common method variance.

With its multidimensional nature, the D-KEFS, in its current form and likely in later editions, stands as an appropriate measure for the development of composite scores and the translation of evidence-based factors into clinical practice. However, to develop these composite scores, a confirmatory factor analysis must first demonstrate that these evidence-based factors explain performance patterns across D-KEFS tests. As noted earlier, three factors of diverse executive functions (i.e., updating, shifting, and inhibition) have garnered substantial empirical support; however, they do not present an exhaustive list of executive functions (Miyake et al., 2000; Miyake & Friedman, 2012). Among the most frequently used terms for executive functions in the literature, five occur most often: planning, working memory, fluency, set-shifting, and inhibition (Packwood et al., 2011). The D-KEFS battery offers a series of tests that tap into many of these established constructs and provides a method for evaluating theoretical executive functions via a norm-referenced and clinically validated battery of tests.

Among the executive function constructs measured by the D-KEFS, those with multiple tests available to serve as indicators in a factor analytic model include fluency, shifting, and inhibition. Both inhibition (i.e., the volitional restriction of a dominant or prepotent response pattern in reaction to a change in task demands) and shifting (i.e., the flexible switching between mental sets) are consistent with those more basic functions described by Miyake and Friedman (2012) and most often reported in latent variable studies on executive function (see Chapter 1), but the third factor consistently evaluated by these researchers, updating (i.e., the monitoring of working memory content, along with the active addition and deletion of said content), is not directly tapped by the D-KEFS test battery. Fluency could serve as a proxy for this construct, because fluency-based tasks may tap into the same cognitive mechanisms as updating. Two mechanisms possibly underlying updating include the "effective gating of information and controlled retrieval from long-term memory" (Miyake & Friedman, 2012, p. 11). Verbal fluency performances require the strategic retrieval of information from long-term memory, involving both working memory and the lexicon (Unsworth, Spillers, & Brewer, 2011; Shao, Janse, Visser, & Meyer, 2014). In turn, within the D-KEFS framework, fluency may serve as the best proxy for updating in current clinical practice, where measures more specific to the updating construct are not commonly administered in clinical settings (e.g., *n*-back; Owen, McMillan, Laird, & Bullmore, 2005).

In addition to these three factors, higher-order dimensions of executive function may be present in the D-KEFS data. Recently published statistical models have shown a common executive function dimension fully explaining the variance in inhibition, while also explaining a significant amount of variance in shifting and updating (Fleming, Heintzelman, & Bartholow, 2016; Friedman et al., 2008; Friedman, Miyake, Robinson, & Hewitt, 2011; Ito et al., 2015). Therefore, the D-KEFS may demonstrate a similar structure, whereby a common executive function explains performances across inhibition, shifting, and fluency. In addition to these three basic factors, additional tasks embedded in the D-KEFS (e.g., the Twenty Questions, Word Context, and Proverbs Tests) may tap into constructs that have not been extensively evaluated in past confirmatory factor analyses (e.g., abstraction, reasoning, and problem solving; Baron, 2004; Delis et al., 2001; Shunk, Davis, & Dean, 2006). These constructs are semantically related to each other, but unrelated to the five terms most frequently used to describe executive functions (i.e., planning, working memory, fluency, set-shifting and inhibition; Packwood et al., 2011) and the D-KEFS offers a method for evaluating whether inhibition, shifting, or fluency are substantially related to these constructs. As a result, these tests may best serve as outcomes in structural models rather than separable factors in measurement models.

The D-KEFS is a widely disseminated, multidimensional clinical instrument tapping into diverse executive-related constructs; and, in turn, the first goal of the proposed study was to use the normative D-KEFS data (Pearson, Inc., 2001) to link research on executive functions with clinical practice by deriving evidence-based factors of executive functions from the tests included in the D-KEFS battery. Achieving this goal involves two research aims that we hope will guide the future development of evidence-based composite scores for executive functions in clinical settings. Specifically, we aimed to (a) derive a three-factor statistical model of executive functions (i.e., inhibition, shifting, and fluency) from the D-KEFS tests, and (b) compare this three-factor model to alternative first-order models (e.g., one-factor, two-factor models) and a bifactor model (Reise, 2012) to determine any support for a common executive function composite.

The second goal of this study was to evaluate how well these factors explained performance on tasks related to abstraction, reasoning, and problem solving using an analytical framework similar to the structural models tested by Miyake and colleagues (2000). To achieve this goal, we aimed to evaluate the relative contribution of each factor at predicting complex task performances through (a) each factor separately serving as a single predictor, (b) all three factors simultaneously serving as predictors, and (c) a common executive function bifactor serving as a sole predictor.

Achieving the first goal will support future linkages between the cognitive research on executive functions and clinical assessment practices, by demonstrating the capacity of the D-KEFS to accurately measure evidence-based factors of executive functions. This will support efforts to develop composite scores using future editions of the D-KEFS based on the current scientific understanding of higher-order cognition. The fulfillment of the second goal will support researchers and clinicians attempting to understand the relationship between inhibition, shifting, and fluency and less commonly evaluated constructs (e.g., abstraction, reasoning, and problem solving).

Method

Participants

Participant data for all analyses came from the D-KEFS normative sample, received with permission from Pearson Incorporated (2001). The norming procedure for the D-KEFS involved a standardized sampling of 1,750 participants ranging from 8 to 89 years of age, with representation of sex, age, race/ethnicity, education, and geographic region consistent with the 2000 U.S. Census data. For further information, the technical manual (Delis et al., 2001) and relevant test reviews of the D-KEFS (Baron, 2004; Homack, Lee, & Riccio, 2005; Shunk et al., 2006) provide a more extensive description of the normative sample. For the current study, an adult sub-sample of the normative group was selected for analyses, with ages ranging from 20 to 49 (N = 425). This age range was selected because the three-factor and nested factor models have been consistently observed in previous studies examining similarly aged samples (Fleming et al., 2016; Ito et al., 2015; Klauer et al., 2010; Miyake et al., 2000). An *a priori* power analysis also estimated the sample size for this age span to be sufficient (i.e., $\hat{\pi} \sim 0.85$) based on the approximate degrees of freedom for the evaluated models (df = ~25; Hancock, 2006). The sample was predominantly White (70.1%), but also included African-American (13.6%) and Hispanic (12.2%) participants. The remaining portion of the sample (4%) came from other ethnic and racial backgrounds. In terms of gender, roughly half of participants identified as male (49.9%) and the remaining identified as female (50.1%).

Materials

The normative data included all primary and optional outcomes for the nine D-KEFS tests (i.e., TMT, Verbal Fluency, Design Fluency, CWIT, Sorting Test, Twenty Questions Test, Word Context Test, Tower Test, Proverb Test). In addition to D-KEFS measures, 358 participants in the selected sub-sample were administered the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), with data from its Vocabulary subtest used in a residualization process described in the control variable subsection of this chapter.

Test assignments to latent constructs. The following subsections describe the tests assigned as indicators to each latent construct specified in the measurement model: inhibition, shifting, and fluency. The final subsection lists the criterion variables that were predicted by these latent variables in structural equation models. Tables 13 and 14 provide reliability estimates for all test scores used in the analysis (as reported in the technical manual; Delis et al., 2001).

Inhibition. The D-KEFS has two tests previously shown to tap into inhibition, the Tower Test and the CWIT. The Tower Test involves participants re-organizing a set of disks on three pegs to match a target design. For each item, the number of moves to completion is weighted, and fewer moves results in a higher score. The primary outcome from the Tower Test is the Total Achievement score, which is a sum of the weighted scores for each trial (i.e., weighted based on number of moves till completion). Although originally designed to test planning abilities, the construct validity of tower test has been questioned (Kafer & Hunter, 1997) and previous structural models have shown that performances on tower tests are related to inhibition (Miyake et al., 2000). The CWIT involves two conditions with an inhibitory component, and both have a primary outcome of time-to-completion: Inhibition and Inhibition/Switching. The Inhibition trial consists of an incongruent Stroop condition, where participants must read the color of the ink and not the word written. The Inhibition/Switching trial requires participants read either the color of the ink or the word written depending on the presentation of the stimulus (i.e., either the word displayed on its own or the word displayed inside a box).

Shifting. Multiple D-KEFS tests include a switching sub-trial that involves shifting for effective performances, with three D-KEFS outcomes selected to load on this factor: the TMT: Number-Letter Switching, Card Sorting Test: Confirmed Correct Sorts, and Design Fluency: Switching – Total Correct. The TMT on the D-KEFS involves a classic Number-Letter Switching trial that requires participants to connect labeled dots in both alphabetical and numerical order, actively switching between number and letter sets. The primary outcome of this test (i.e., time-to-completion) served as an indicator on the shifting factor in the measurement model. The second shifting indicator derived from the Card Sorting Test, where participants arrange cards based on verbal and visual-spatial characteristics. Participants are not told how to sort the cards, and must shift from previous sorting rules to new rules in order to attain a greater number of accurate sorts. The Card Sorting Test outcome used in the current factor model (i.e., Confirmed Correct Sorts) has correlated significantly with performance on the Wisconsin Card Sorting Test (WCST; i.e., $r = 0.59$; Delis et al., 2001), which requires shifting for effective task performance (Miyake et al., 2000). The last indicator for the shifting factor came from the Switching trial of the Design Fluency test where clients had to draw novel abstract designs, while switching between connecting black and empty dots. The total number of unique designs accurately drawn served as the outcome for this task.

Fluency. Two D-KEFS tests target fluency through verbal and visual output, with three outcomes from these measures serving as indicators on the fluency factor: two outcomes from the Verbal Fluency test (i.e., Letter Fluency – Total Correct and Category Fluency – Total Correct) and one outcome from the Design Fluency test (i.e., Filled + Empty Dots – Total Correct). The D-KEFS includes a test of Verbal Fluency that involves both phonemic and semantic trials. The phonemic trial, Letter Fluency, involves three one-minute sub-trials, where participants attempt to name as many words as possible that begin with a specific letter. The semantic trial, Category Fluency, consists of two sub-trials, both of which require participants to name as many words as possible that belong to a specific category. For both the phonemic and semantic trials, a total score comes from summing the words generated across all sub-trials, with these outcomes serving as indicators in the measurement model. The third fluency indicator came from the Design Fluency test, which involves two trials where participants connected empty or filled dots to draw a series of novel abstract designs. The total unique designs drawn across these two trials served as the final indicator.

Control variables. Indicators deriving from the same tests were included on separate factors, which led to high correlations between indicators due to shared method variance rather than true construct variance. To control for this method variance, the original approach was to include orthogonal control factors in the model that involved the co-loading of indicators onto both an executive function factor (e.g., verbal fluency tests loading onto the fluency factor) and a control factor (e.g., verbal fluency tests loading onto a vocabulary factor along with the WASI Vocabulary score). However, some control factors and executive function factors shared two out of three of their indicators, resulting in high correlations between these factors, and a model that forced these factors to be uncorrelated did not properly converge.

As an alternative approach, executive function indicators were residualized of method variance using ordinary least squares regression models, making their scores orthogonal to those of the control variables. In order to control for processing speed, the two CWIT scores loading on the inhibition factor were made orthogonal to the summed performance on the CWIT Word Reading and Color Naming trials. Also to control for processing speed, the TMT Number-Letter Switching trial score that loaded on the shifting factor was made orthogonal to the summed performance on the TMT Number and Letter Sequencing trials. The Design Fluency Switching score that loaded on the shifting factor was made orthogonal to the Design Fluency score that loaded on the fluency factor. Lastly, the Verbal Fluency scores were made orthogonal to the WASI Vocabulary subtest, controlling for the impact of language functioning on these outcomes (Unsworth, Spillers, & Brewer et al., 2011; Shao et al., 2014).

Criterion variables. Three measures from the D-KEFS (i.e., Twenty Questions Test, Word Context Test and Proverb Test) will serve as criterion variables in structural models, predicted by the latent variables included in the measurement model. These D-KEFS tests purportedly require abstraction, reasoning, and problem solving for effective performance (Baron, 2004; Delis et al., 2001; Shunk et al., 2006), which stand as semantically unique constructs from the latent variables included in the measurement model (Packwood et al., 2011).

Statistical Analysis

The D-KEFS normative data were received by Pearson Inc. already age-corrected and standardized for all D-KEFS variables (*M* = 10, *SD* = 3) and WASI variables (*M* = 50, *SD* = 10), with a higher value indicating a greater performance. All structural equation modeling was conducted in MPlus v.7.3 (Muthén & Muthén, 2014), using maximum likelihood to estimate model parameters based on the covariance matrix of the selected D-KEFS outcomes. A full information maximum likelihood estimation approach was used to handle missing data, because this approach provides an unbiased technique for the analytical treatment of missing data within the context of structural equation modeling (Enders & Bandalos, 2001). Missing data was assumed to be missing completely at random. Appendix C provides the MPlus syntax for all models evaluated.

Fit indices. All models were evaluated in relation to statistical fit, with optimally fitting models selected to guide interpretation. The reported fit indices include the χ^2 Test of Model Fit, the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI). The χ^2 test involves the comparison between the observed covariances and the predicted covariance of the specified model, with an insignificant value indicative of good model fit (Hoyle & Panter, 1995). Notably, with large sample size, the χ^2 test has increased power to reject the null hypothesis (Tanaka, 1987), and alternative fit indices will serve as primary metrics for interpretation considering the large normative sample included in the current analyses. The AIC represents a modification of the χ^2 statistic that penalizes for model complexity, with values closer to zero representing better fit (Akaike, 1974). The BIC is closely related to the AIC, and can also be interpreted as a value closer to zero providing better fit (Schwarz, 1978).

The CFI and TLI provide alternative assessments of fit (Bentler, 1990), with values closer to 1.0 considered desirable (Cheung & Rensvold, 2002). The RMSEA also provides an alternative fit index, approximating lack of model fit compared to the saturated model, with values closer to zero being desirable. Past researchers have posited optimal fit cutoffs of $\geq .95$ for the CFI and TLI and ≤ 0.06 for the RMSEA (Hu & Bentler, 1999). Some alternative recommendations provide greater leniency on these cutoffs, including a lower-bound cutoff of $\geq .90$ for the CFI (Bentler & Bonett, 1980) and a higher-bound cutoff of $\leq .07$ for the RMSEA (Steiger, 2007).

For model comparison, certain fit indices have metrics for evaluating significant changes in model fit. The CFI difference (ΔCFI) and χ^2 difference ($\Delta\chi^2$) tests provide metrics for model comparison. The $\Delta\chi^2$ test allows for nested model comparisons, with a significant value indicating a significant reduction in model fit. Because $\Delta\chi^2$ has sensitivity to large sample size and the models evaluated in the current study were non-nested, the ΔCFI will serve as the primary metric for model comparison. The recommend cutoff of $\Delta\text{CFI} \geq .01$ was used to compare models in the proposed analyses (Cheung & Rensvold, 2002). The ΔBIC can also be used in model comparison, where a BIC value closer to zero is considered preferable. Typical cutoffs for the ΔBIC include 0-2 for weak, 2-6 for positive, 6-10 for strong, and >10 for very strong evidence of one model being preferable over another (Raftery, 1995).

Measurement models. The hypothesized three-factor model (i.e., inhibition, shifting, and fluency) was evaluated first for statistical fit, and alternative measurement models were evaluated thereafter. These alternative models included a one-factor model and a set of three possible two-factor models that merged two of the first-order factors (i.e., inhibition = fluency; fluency = shifting; inhibition = shifting). A bifactor model was also evaluated, where indicators co-loaded on their specific factor (i.e., inhibition, shifting, fluency) and a general factor (i.e., common executive function). Aligning with past research on bifactor models of executive function tasks (Miyake & Friedman, 2012), an incomplete bifactor model (Chen, West, & Sousa, 2006) was also evaluated, where a specific inhibition factor was not included in the model. For all models evaluated, 5,000 bootstrapped samples were derived to calculate how often among those samples the factor model "properly converged," meaning the model both converged and provided an admissible solution (e.g., no correlation above 1.0, no negative residual variances, etc.).

Reliability. The factors included in each model were evaluated for their reliability using omega (ω) as an estimate, which represents the ratio of true-score variance to the total variance among indicators loading onto a given factor (McDonald, 1999). In the context of the bifactor model, the reliability of the general and specific factors was calculated using omega-hierarchical (ω_H) and omega-subscale values (ω_S), which provide estimates of variance accounted for by the bifactor and the specific factors (i.e., inhibition, shifting, fluency), respectively (Reise, 2012). In a context where a battery of tests shows high unidimensionality, ω_H would be high, while the ω_S values would be much lower (Canivez, 2016). There are no commonly used cutoff scores to guide interpretation of ω values, but the values can be interpreted as the amount of variance in the indicators attributable to the factor and not error. Because executive function tests suffer from task impurity, which would result in greater error variance, relatively low ω values are likely common in the context of executive function measurement.

Structural model. The best-fitting first-order factor model – and bifactor model if it showed acceptable fit – were treated as the accepted models, with their latent variable(s) used as predictors in structural models. In a series of models, either diverse (e.g., inhibition, shifting, fluency) or unitary (i.e., common executive function) factor(s) predicted performances on three complex tasks of executive functions (i.e., the Twenty Questions Test, the Word Context Test, and the Proverb Test). The path coefficient from each factor to each complex task was evaluated for its unique significance, along with the R^2 value for each model to determine the amount of variance accounted for in each task.

Results

The descriptive statistics for all variables included in the measurement and structural models are provided in Table 15. Skewness and kurtosis were within normal limits, indicating an approximate univariate normal distribution for each variable. Multivariate normality was evaluated via Mahalanobis' distance, which identified one multivariate outlier. Results did not change with the removal of this multivariate outlier; and, in turn, this participant's data was included in all analyses. The bivariate correlations for all measures are provided in Table 16. Correlations between inhibition tasks were all significant and ranged from r = 0.136 to 0.467. Correlations between shifting tasks ranged from 0.032 to 0.250, with the lowest – and only non-significant – correlation occurring between the TMT and the Design Fluency Switch trial. The correlations between fluency tasks were all significant, ranging from 0.180 to 0.443. The highest correlations occurred between variables deriving from the same tests: CWIT outcomes correlated at 0.467 and Verbal Fluency outcomes correlated at 0.443.

Measurement Models

Table 17 provides the fit indices for each model evaluated. The hypothesized three-factor model (i.e., inhibition, shifting, and fluency) did not meet full criteria for adequate fit (CFI = 0.871; RMSEA = 0.066). Modification indices indicated that a correlation between the errors of the two Verbal Fluency tasks would significantly improve model fit. Considering the shared method variance between these tasks, this correlation was added to the model, which resulted in the expected increase in model fit. The three-factor model with correlated errors for Verbal Fluency scores had good statistical fit (CFI = 0.938; RMSEA = 0.047). The fit for the three-factor model surpassed the fit of the unidimensional model (CFI = 0.838; RMSEA = 0.072) and all models with merged factors. Although fit was much greater for the three-factor model in comparison to models that merged inhibition with shifting (CFI = 0.871; RMSEA = 0.065) or fluency (CFI = 0.854; RMSEA = 0.069), the model that merged fluency and shifting showed good fit (CFI = 0.929; RMSEA = 0.048) that was not significantly different than the fit of the three-factor model (ΔCFI = 0.009), but came extremely close to the ΔCFI $\geq$ .01 threshold for a significant difference. The ΔBIC between these two models was 6.803 with a preference for the merged factor model; however, all other fit indices showed better fit for the three-factor model. In turn, the three-factor model was the accepted first-order factor model. Based on the bootstrapping analysis, this model properly converged at a frequency of 97.98%, with convergence frequencies for other models presented in Table 17. Reliability estimates were calculated for each latent variable included in the three-factor model, resulting in ω values of 0.59 for inhibition, 0.38 for shifting, and 0.38 for fluency.

The bifactor model was also evaluated and showed the best fit based on some indices (CFI = 0.977; RMSEA = 0.032). It was the only model to provide a non-significant χ^2 value (p = 0.1006) despite the large sample size, but it had a ΔBIC of 11.682 when compared to the three-factor model. The bifactor model also properly converged among only 65.86% of the 5,000 bootstrapped samples. The bifactor model without a specific inhibition factor did not converge. Reliability estimates were calculated for the general and specific factors: ω_H came to 0.49, while the ω_S values were 0.39 for inhibition, 0.03 for shifting, and 0.48 for fluency. Despite its low rate of convergence, the bifactor model was used in structural models in addition to the three-factor model, considering its excellent fit based on some indices and the parsimony of interpreting the general factor.

Figures 9 and 10 provide visual representations of the three-factor and bifactor models, respectively. Each figure includes loadings and errors. Only the figure for the three-factor model provides inter-factor correlations, because the factors were orthogonal in the bifactor model. All loadings were significant for the three-factor model; however, some loadings in the bifactor model were non-significant. These non-significant loadings included the TMT on the shifting-specific factor and Verbal Fluency phonemic trial on the bifactor. Although non-significant, their removal did not improve statistical fit and they were included in all structural models.

Structural Models

Two series of structural models were conducted, involving the prediction of three tests (i.e., Twenty Questions Test, the Word Context Test, and the Proverb Test) using latent dimensions of executive function. The first series used the diverse executive function factors included in the accepted first-order three factor model to predict the tasks, and the second series used the common executive function factor from the bifactor model to predict the tasks. These two approaches served to evaluate (a) which diverse executive function best predicted performance on the criterion tasks, and (b) whether the common executive function or the diverse set of executive functions accounted for more variance in the tasks. The first series included three models in which all three first-order factors simultaneously predicted each task, and follow-up models in which each diverse factor predicted each task one at a time. The second series consisted of three models, where the common executive function factor predicted each task separately.

Table 18 provides results for all structural models. The bifactor model significantly predicted all complex tasks, accounting for a significant amount of variance in each outcome. The models involving the simultaneous prediction of each task by inhibition, shifting, and fluency accounted for a significant amount of variance in all tasks except for the Twenty Questions Test, although the R^2 value for this model approached significance (p=0.057). Notably, in all models involving simultaneous prediction, shifting was the only uniquely significant predictor. In models where only inhibition predicted the criterion, inhibition was a significant predictor; and in models where only shifting predicted the criterion, shifting was also a significant predictor.

However, in models where only fluency predicted the criterion, the models did not properly converge due to a correlation above 1.0 between fluency and shifting.

Discussion

A series of confirmatory factor analyses on a subset of D-KEFS tests identified a three-factor solution as the best-fitting measurement model. This model included three executive functions posited by past researchers (Miyake et al., 2000; Kramer et al., 2014; Packwood et al., 2011): inhibition, shifting, and fluency. Inter-factor correlations were high between shifting and inhibition (r = 0.591) along with shifting and fluency (r = 0.696), although much lower between inhibition and fluency (r = 0.360). The high correlations between shifting and the other factors likely derive from the close relationship between shifting and the common executive function dimension. A bifactor model showed that a general factor, representative of common executive function, explained nearly all variance in the shifting indicators, leaving virtually no unique variance for a shifting-specific factor (ω_S = 0.03). Unlike previous measurement models showing inhibition as fully explained by a common executive function factor (Fleming et al., 2016; Friedman et al., 2016; Ito et al., 2015), a bifactor model without inhibition failed to converge for the D-KEFS battery. Although the bifactor model fit the data better than the three-factor model, the model properly converged among only 65.86% of 5,000 bootstrapped samples; and while this model was informative for explicating the relationship between shifting and a general executive function dimension, the replicability of this model remains questionable.

A robust general factor does not come as a surprise considering the high inter-factor correlations often observed by previous confirmatory factor analyses on executive function test batteries and the many studies supporting a nested factor model (Fleming et al., 2016; Friedman et al., 2008, 2016; Ito et al., 2015; Klauer et al., 2010; Miyake et al., 2000). While a robust general bifactor is consistent with previous research, the relationship between shifting and a common executive function dimension was novel, and this relationship can be interpreted in multiple ways. A first interpretation could consider these results as evidence for a uniquely strong relationship between shifting and common executive function. Past researchers have posited that shifting arises last in executive function development and successful shifting performance may require inhibitory and working memory abilities (Best & Miller, 2010). The findings of Chapter 1 also evidenced shifting as arising later in development. This conceptualization potentially indicates shifting as higher-order, which aligns with previous claims that – in early development – shifting requires the establishment of more basic executive abilities prior to successful performance (Müller & Kerns, 2015). In terms of implicated brain regions, there is evidence for a superordinate fronto-cingulo-parietal network being activated during tasks of diverse executive functions (i.e., inhibition, working memory, and flexibility; Niendam et al., 2012), which supports a common executive function ability but does not indicate one ability as more central than another. In terms of past research, all studies examining the structure of executive function among adults have identified shifting as a separable factor, and many have supported inhibition, rather than shifting, as more closely related to the common executive function bifactor (Fleming et al., 2016; Fournier-Vicente et al., 2008; Ito et al., 2015; Klauer et al., 2010; Miyake et al., 2000). However, Chapter 1 did not indicate any model as unequivocally preferred based on previous confirmatory factor analyses of executive function test batteries, and a substitute model with a central shifting ability may be an accurate representation of the true nature of executive function. Alternatively, other possible interpretations for the near perfect relationship between common executive function and shifting may consider the nature of the tests assigned to their measurement.

The tests that loaded onto the shifting factor in the accepted model conceptually and empirically align with general executive function abilities. An outcome from the Sorting Test loaded highest onto the shifting factor, and a previous analysis has shown a strong relationship between the Sorting Test and performances on the commonly administered WCST (i.e., r = 0.59; Delis et al., 2001). Although the WCST does relate to shifting ability (Miyake et al., 2000), the latent factor underlying WCST performance has been conceptualized as a general executive function factor due to the complexity of the task (Greve, Stickle, Love, Bianchini, & Stanford, 2005). The TMT had the second highest loading on the shifting factor, but this loading was non-significant in the bifactor model, indicating that performance on the TMT was highly related to common executive function. A recent analysis of the construct validity of the TMT found that both working memory and task-switching explained performance of the switching condition (Sanchez-Cubillo et al., 2009). In turn, the tasks used as indicators for shifting likely tapped into shifting, but also other executive-related constructs, making the variance within this factor more general than specific.

In comparison to the shifting factor, the more modest relationship between the general factor and the specific inhibition and fluency factors may also derive from the nature of the tests assigned to these factors. Method variance influenced the findings of a previous factor analysis on the D-KEFS (Latzman & Markon, 2010), and has potentially impacted the current findings as well. In the context of the bifactor model, the specific inhibition (ω_S = 0.39) and fluency (ω_S = 0.48) factors had far more unique variance than the shifting factor (ω_S = 0.03); however, the co-loadings for some of their respective indicators decreased, suggesting the bifactor explained some variance in those indicators. Specifically, the inhibition-specific loading for the Tower Test decreased from the three-factor (λ = 0.289) to the bifactor model (λ = 0.144); and the fluency-specific loading for Design Fluency decreased from the three-factor (λ = 0.627) to the bifactor model (λ = 0.187) as well. The indicators loading the highest on the specific factors for these constructs were related based on method variance. Two verbal fluency scores loaded highly on the fluency-specific factor (i.e., λ = 0.572 and 0.730), and two CWIT scores loaded highly on the inhibition-specific factor (i.e., λ = 0.466 and 0.745). In turn, the strong relationship between shifting and the general factor could have resulted from a lack of common method variance between shifting indicators in comparison to the other specific factors, because shifting was the only diverse factor with indicators from three different D-KEFS tests.

Aside from examining the latent structure of the D-KEFS, the current study also evaluated the relationship between executive function factors and tests purported to measure the constructs of abstraction, reasoning, and problem solving. In a series of models, the bifactor or the three diverse factors (i.e., inhibition, shifting, and fluency) predicted a set of complex executive function tasks not often evaluated in latent variable research (i.e., the Twenty Questions, Word Context and Proverb Tests). The bifactor significantly predicted performance on all of these tasks. In comparison, when all diverse factors predicted these tasks simultaneously, shifting was the only factor to significantly predict complex task performance. In models including just a single path from one diverse factor to the criterion variable, shifting and inhibition predicted all tasks significantly on their own; however, when fluency served as the sole predictor, the models did not converge due to high co-linearity between shifting and fluency. These two factors were highly related to one another in the first-order measurement model (r = 0.696), and an alternative measurement model that merged these two factors showed relatively good fit, although multiple indices suggested its fit was worse than the three-factor model. The models involving shifting as the sole predictor yielded higher R^2 values than the models with inhibition or the bifactor as sole predictors. Models including three paths from inhibition, shifting, and fluency had higher absolute R^2 values than any single path models, although the R^2 for the Twenty Questions Test was non-significant. These findings indicate that, for the three-factor model, shifting explained the most variance in the criterion outcomes, likely due to its association with the common executive function factor. In turn, the constructs measured by these three criterion tasks (i.e., the Twenty Questions, Word Context, and Proverb Tests) are closely related to either shifting or general executive function ability.

The results discussed above achieved the primary aims of this study: (a) determining the latent structure for the D-KEFS test battery and (b) determining which constructs are most closely related to performances on a set of criterions tasks. An overarching goal that guided the pursuit of these two aims was the linkage between the current findings and clinical practice; however, some limitations to the current design disallow the immediate clinical application of these results. The original statistical approach aimed to include control factors on which variables that shared method variance co-loaded (e.g., a vocabulary factor accounting for common method variance in Verbal Fluency outcomes and a speed factor accounting for common method variance in time-based outcomes). These control factors would be orthogonal to the executive function factors, making the inhibition, shifting, and fluency factors pure of method variance not attributable to the constructs of interest. However, any model including these co-loadings would either not converge or provide an inadmissible solution, potentially due to the high co-linearity between the executive function and control factors. In turn, the residualization of variables was conducted prior to the confirmatory factor analysis. This approach allowed for variables to be free of common method variance attributable to speed and language ability. However, the D-KEFS does not provide normative data for residualized scores, and summed scaled scores from the D-KEFS would not fully correspond to the factors identified in the accepted model.

Another notable limitation of the current study was the inclusion of scores from the same D-KEFS test on either the same or different factors. The D-KEFS does offer different conditions within each test that tap into different constructs, but these conditions often rely on a common method of measurement. Although the residualization procedure attempted to control for method variance, the use of multiple indicators from the same tests led to shared method variance between CWIT and Verbal Fluency scores infiltrating the inhibition and fluency factor scores, respectively. As well, the two orthogonalized Design Fluency scores loaded on the shifting and fluency factors; and while this approach controlled for shared method variance between indicators, it may have biased the model towards a multidimensional solution by orthogonalizing indicators assigned to separate, but correlated factors.

Although the residualization process prevents the calculation of composite scores, the latent structure identified herein offers an empirical framework through which clinicians can produce a valid interpretation of their assessment findings. The modeling results demonstrate the advantages of the D-KEFS in its capacity to tap into three latent constructs posited by previous researchers on executive functions along with more global executive-related abilities (Packwood et al., 2011). The findings also identify potential considerations for future editions of the D-KEFS, where the latent structures of the battery may be hypothesized and evaluated during test development, potentially providing composite scores for use in clinical practice.

By design, the D-KEFS does not provide composite scores (Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003), but instead offers a wide array of scores evaluating many proposed dimensions of executive function. Delis and colleagues questioned the utility of composite scores among clinical samples, because unitary constructs dissociate among clinical populations. Despite these concerns, past researchers have provided methods for the calculation of an atheoretical global composite score using the D-KEFS (Crawford, Garthwaite, Sutherland, & Borland, 2011); although the scores included in this calculation do not align with the current findings or any previous research on executive functions. Efforts to provide D-KEFS composite scores in future editions of the test battery would benefit from considering the theoretical structure of executive functions in their calculation, ensuring the highest possible level of reliability and construct validity.

Since the first confirmatory factor analysis on executive functions, researchers have recommended the evaluation of latent factors in clinical settings (Miyake, Emerson,

& Friedman, 2000); however, none of the few composite scores for executive functions available through performance-based clinical neuropsychological instruments correspond to the diverse executive functions identified by previous researchers (e.g., the Neuropsychological Assessment Battery Executive Functions Module, White & Stern, 2003; the Test of Executive Control, Isquith et al., 2009). Individual tests examining executive functions have been criticized for low reliability (Schmidt, 2003) and poor ecological validity (Barkley, 2012; Chaytor, Schmitter-Edgecombe, & Burr, 2006). Considering the improved purity of factor estimates of executive functions (Miyake & Friedman, 2012), the use of composite scores in clinical practice, analogous to evidence-based factors, may provide a more robust estimate of executive functions at the individual level.

Without composite scores, the current assessment practices for executive functions appear inconsistent compared to the assessment practices for other cognitive domains. Composite scores have become particularly common in clinical practice due to their ability to simplify psychometric interpretation by summarizing performances across tests. They have become almost ubiquitous in the assessment of intellectual functioning, where composite scores of intelligence have strong reliability and predictive validity (Kamphaus, 2005; Lubinski, 2004; Sternberg, Grigorenko, & Bundy, 2001). Although composite scores allow clinicians to simplify complex test profiles, they are unfortunately unavailable for the D-KEFS.

Given that the current findings do not directly translate into clinical practice and that the D-KEFS was not constructed with the intention of providing composite scores (Delis et al., 2003), an alternative method to guide a multivariate assessment of executive functions would aid clinicians attempting to make sense of complex test profiles. Multivariate base rates quantify the prevalence of low scores among the standardization sample for a test battery, allowing clinicians to take multiple scaled scores and determine the normal frequency at which a client may fall under a low score cutoff when administered multiple tests (e.g., ≤16 percentile, ≤5th percentile, etc.). These base rates have become available for many popular test batteries (e.g., Wechsler Adult Intelligence Scale, Fourth Edition; Wechsler Memory Scale, Fourth Edition; Brooks, Iverson, & Holdnack, 2013a), and they do not require the calculation of composite scores to summarize a test profile. They can also improve the diagnostic accuracy of a neuropsychological assessment (Brooks, Iverson, Feldman, & Holdnack, 2009; Brooks, Iverson, & White, 2007). Past researchers have provided multivariate base rates for the D-KEFS (Crawford et al., 2011); but these researchers did not have access to the D-KEFS normative data, and they did not provide stratifications for relevant demographic characteristics known to affect low score prevalence (e.g., intelligence, education; Brooks et al., 2013a). While D-KEFS composite scores may be on the horizon in future test editions, they are not yet available to neuropsychologists, and multivariate base rates offer a quantitative resource to interpret complex test profiles where composite scores are not yet available. The aim of the next chapter was to calculate multivariate base rates from the D-KEFS normative data with stratifications for age, intelligence, and education, supporting a multivariate interpretation of the most commonly administered executive function test battery in clinical practice.

Chapter 3: Multivariate Base Rates of Low Scores on the Delis-Kaplan Executive Function System

Abstract

Multivariate base rates allow for the *simultaneous* interpretation of multiple test scores, quantifying the normal frequency of low scores on a test battery. Multivariate base rates for the Delis-Kaplan Executive Function System (D-KEFS) would support clinicians interpreting performance patterns across multiple tests of executive functions, reducing the likelihood of misdiagnosing executive dysfunction through the over-interpretation of univariate low scores. The D-KEFS consists of 9 tests with 16 Total Achievement scores (i.e., primary indicators of executive function ability). Stratified by age, education, and intelligence, multivariate base rates were derived for the full D-KEFS and two brief batteries using the adolescent and adult portion of the normative sample (i.e., ages 16-89; N = 1,050). Summary tables provide multivariate base rates for the full and brief D-KEFS batteries. Base rates are provided for separate age spans (i.e., 16-89, 16-69, and 60-89) stratified by years of education and intelligence. The base rates are presented as cumulative percentages of the normative sample obtaining one or more low scores. Low scores occurred commonly among the D-KEFS normative sample. Their frequency increased with the number of scores interpreted and decreased with the use of lower clinical cutoff scores. Low scores occurred more frequently among individuals with lower intelligence and less education. The prevalence of impaired scores appears greater for the D-KEFS than other commonly administered test batteries (e.g., WAIS-IV, WMS-IV), suggesting a greater risk of misdiagnosing cognitive impairment when evaluating executive functions in comparison to other domains.

Introduction

The accurate measurement of executive functions has been challenging for clinicians and researchers (Chan, Shum, Toulopoulou, & Chen, 2008), potentially due to the difficulty of defining the construct (Jurado & Rosseli, 2007). Since the first use of the term three decades ago (Lezak, 1982), a multitude of definitions for executive functions have emerged, with the majority describing the construct as multidimensional (Barkley, 2012; Baggetta & Alexander, 2016). Researchers have proposed a wide array of cognitive sub-components falling under the umbrella term of executive functions (Packwood, Hodgetts, & Tremblay, 2011). Accounting for these many component abilities, a more recent definition of executive functions has described the construct not just as a set of self-regulatory processes, but rather an interactive system that works collaboratively to produce volitional, goal-oriented behavior (Jurado & Rosseli, 2007). In turn, for an accurate evaluation of executive functions, clinicians conduct multidimensional assessments incorporating multiple tests that examine diverse executive functions in aggregate, because a single test assessment and/or univariate interpretation of executive functions likely misrepresents the true nature of higher-order cognitive functions (Baggetta & Alexander, 2016; Miyake, Emerson, & Friedman, 2000). Repeated surveying of clinicians has consistently indicated that the assessment of executive functions commonly relies on the use of multiple standalone tests (Rabin, Barr, & Burton, 2005; Rabin, Paolillo, & Barr, 2016), with limited quantitative resources to form a multivariate interpretation of test findings.

The assessment of executive functions has evolved from single tests to the development of increasingly popular broadband batteries of co-normed tests evaluating the domain (Jurado & Rosseli, 2007), with the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001) representing the most commonly administered of these batteries in current clinical practice and research (Baggetta & Alexander, 2016; Rabin et al., 2016). Past practices for the assessment of executive functions involved the administration of multiple narrowband measures with separate normative samples. Clinical inferences and interpretations varied depending on the normative sample selected by the clinician (Kalechstein, van Gorp, & Rapport, 1998), and a co-normed battery of multiple tests represented a significant step forward (Homack, Lee, & Riccio, 2005). However, although using a co-normed battery has distinct advantages, clinicians and researchers still face the challenge of how to interpret multiple test scores from a single client or research participant. Chapter 2 aimed to identify a factor model for the D-KEFS that would guide the development of evidence-based composite scores to summarize performances across tests of common constructs; however, the findings were not directly translatable to clinical practice, and an alternative approach is necessary to support clinicians attempting to make a multivariate interpretation of a test profile.

With no quantitative method for interpreting scores in aggregate, clinicians often make multiple univariate interpretations when assessing executive functions, rather than forming a cohesive multivariate interpretation of their test findings. Univariate interpretations fail to account for the normal occurrence of low scores when multiple tests are administered and interpreted (Binder, Iverson, & Brooks, 2009; Brooks, Iverson, &

Holdnack, 2013a). Low scores occur quite commonly among healthy individuals completing any battery of tests, and vary as a function of the intelligence and demographic characteristics of participants and clients (Binder et al., 2009; Brooks & Iverson, 2012; Brooks et al., 2013a; Iverson & Brooks, 2011; Schretlen, Testa, Winicki, Pearlson, & Gordon, 2008). Past researchers have derived multivariate base rates of low score prevalence for multiple neuropsychological measures (e.g., Halstead-Reitan, Binder et al., 2009; NEPSY-II, Brooks, Sherman, & Iverson, 2010; the Wechsler Adult Intelligence Scale, Fourth Edition [WAIS-IV] and Wechsler Memory Scale, Fourth Edition [WMS-IV], Brooks, Iverson, & Holdnack, 2013), including measures assessing executive functions (e.g., Test of Verbal Conceptualization and Fluency; Brooks et al., 2013b; Brooks, Iverson, Lanting, Horton, & Reynolds, 2012). These base rates quantify the number of individuals within the normative sample with one or more low scores based on commonly used clinical cutoffs, and their use guards against over-interpretation of a single low test score within a battery (Brooks & Iverson, 2012; Brooks et al., 2013a).

Considering the multidimensional nature of executive functions and the array of scores available, clinicians need resources to interpret multivariate data with greater ease. Multivariate base rates assist in quantitatively evaluating performance trends across multiple tests *simultaneously*, understanding how commonly a low score would occur for an individual administered multiple measures within the context of a clinical assessment. This technique moves away from a univariate test-by-test interpretation that may result in diagnostic errors (Brooks, Iverson, Feldman, & Holdnack, 2009; Brooks, Iverson, Holdnack, & Feldman, 2008; Brooks, Iverson, & White, 2007). Past researchers have produced multivariate base rates for the D-KEFS (Crawford, Garthwaite, Sutherland, &

Borland, 2011); but in the context of this past derivation, the frequencies of low scores were estimated from summary data published in the D-KEFS technical manual, with no stratifications based on client characteristics, such as education level or intellectual functioning, both of which are known to influence the frequency of low scores (Binder et al., 2009; Brooks et al., 2013a; Iverson & Brooks, 2011).

Through use of the D-KEFS normative data, the current study aimed to refine these past efforts and calculate multivariate base rates for D-KEFS scores stratified by age, intelligence, and education, hypothesizing that the low score frequencies observed within the D-KEFS will align with results from past studies using other test batteries (Crawford, Garthwaite, & Gault, 2007; Binder et al., 2009; Brooks et al., 2010, 2012, 2013a; Brooks, Holdnack, & Iverson, 2011). Specifically, low scores on the D-KEFS will occur commonly among healthy individuals within the normative sample; and the prevalence of these low scores will increase with the interpretation of more tests, and decrease with the use of lower cutoffs (Brooks & Iverson, 2012; Brooks et al., 2013a; Schretlen et al., 2008). Further, as observed in past base rate derivations, those individuals with greater education and higher levels of intelligence will obtain fewer low D-KEFS scores (Brooks et al., 2011, 2013a; Schretlen et al., 2008).

In addition to furthering the field of multivariate base rate research, a goal of this study is to produce meaningful data relevant to the everyday practice of neuropsychologists. Because most clinicians likely take a flexible approach to the administration of the D-KEFS (Baron, 2004; Homack et al., 2005; Shunk, Davis, & Dean, 2006; Swanson, 2005), the multivariate base rates provided herein were calculated for three permutations of the D-KEFS test battery, including the full battery and two brief batteries composed of the most commonly administered tests of executive functions (Rabin et al., 2016). The low score frequencies derived for the full and brief batteries will facilitate the multivariate assessment of executive functions in neuropsychological practice, assisting clinicians with the *simultaneous* interpretation of multiple D-KEFS scores.

Method

Participants

The D-KEFS normative sample, received with permission from Pearson Incorporated (2001), consisted of a nationally representative sample of 1,750 participants between the ages of 8 and 89. The data were received already standardized for all variables, with all participant scores age-corrected with a mean of 10 and a standard deviation of 3. The representation of sex, age, race/ethnicity, education, and geographic region for the sample closely matched that of the 2000 U.S. Census data. For the current study, only participants between the ages of 16 and 89 were included in the calculation of multivariate base rates, which corresponded closely to the normative age span for the WAIS-IV (i.e., ages 16-90). Multivariate base rates for the pediatric sample will be provided in a future publication.

The normative sample for participants ages 16-89 (N = 1,050) consisted of roughly equal male (47.8%) and female (52.2%) participants across the following eight age groups with norm group size in parentheses: ages 16-19 (175), 20-29 (175), 30-39 (150), 40-49 (100), 50-59 (100), 60-69 (125), 70-79 (125), 80-89 (100). The racial/ethnic makeup was predominantly White (79.2%), with African American (10.6%), Hispanic (8.1%), and participants of other races/ethnicities (2.1%) comprising the remainder of the sample. Participants dispersed across multiple educational levels, including less than or equal to 8 years (5.7%), 9-11 years (12.2%), 12 years (34.3%), 13-15 years (26.7%), and 16 or more years (21.1%). The D-KEFS technical manual (Delis et al., 2001) provides additional information regarding the full normative sample. A sub-sample of the participants included in this study (n = 823) completed the Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999). This sub-sample was used for the calculation of multivariate base rates stratified by intellectual ability level.

Materials

The D-KEFS battery contains a series of nine tests explained in detail in previous test reviews (Baron, 2004; Homack et al., 2005; Shunk et al., 2006; Swanson, 2005). Each of these nine tests has one or more Total Achievement scores listed in the D-KEFS technical manual (Delis et al., 2001), with these scores representing the primary indicators of executive function ability for each test. Table 19 provides a brief summary of all Total Achievement scores derived from the D-KEFS.

As stated earlier, the multivariate base rates were calculated based on the frequency of low Total Achievement scores for the full battery and two brief batteries. The two brief batteries consist of three tests (i.e., Trail Making Test, Verbal Fluency Test, and Color-Word Interference Test) and four tests (i.e., Trail Making Test, Verbal Fluency Test, and Color-Word Interference Test, and Tower Test), selected based on the frequency of their administration in clinical practice (Rabin et al., 2016). The full test battery includes 16 Total Achievement scores, the four-test battery includes 9 Total Achievement scores, and the three-test battery includes 7 Total Achievement scores. As noted earlier, intelligence level was used to stratify the multivariate base rates, which was estimated for a subsample of participants through the WASI. The WASI consists of four subtests that are combined to estimate Full Scale IQ (i.e., Vocabulary, Similarities, Block Design, Matrix Reasoning).

Statistical Analysis

Multivariate base rates of low scores were calculated for three age spans corresponding to the full sample (ages 16-89), adult sample (ages 16-69), and older adult sample (ages 60-89). Because the D-KEFS normative data was received stratified by 10-year age bands (e.g., 60-69), it was not possible to stratify the groups at mid-decade (e.g., 65 years), which resulted in a slight overlap between the adult and older adult samples. These age spans were selected because they corresponded as closely as possible with the age spans for the two forms of the WMS-IV Adult (ages 16-69) and Older Adult (ages 65-90) Record Forms, for which multivariate base rates are already available (Brooks et al., 2013a). Only cases with no missing test data were included in the calculation of the base rates, which explains the fluctuations in sample sizes across different test batteries within each sample (see Tables 20-28 for sample sizes). The multivariate base rates were calculated as cumulative percentages, representing the percent of participants with one or more low scores based on commonly used clinical cutoffs, including ≤25th percentile, ≤16th percentile, ≤9th percentile, ≤5th percentile, and ≤2nd percentile. The base rates were further stratified based on education (i.e., less than or equal to 8 years, 9-11 years, 12 years, 13-15 years, and 16 or more years) and intelligence (i.e., WASI FSIQ ≤89, 90-99, 100-109, and 110+).

Results

The prevalence rates of low D-KEFS Total Achievement scores, based on ≤25th percentile, ≤16th percentile, ≤9th percentile, ≤5th percentile, and ≤2nd percentile cutoffs, for the total sample (i.e., ages 16-89) are provided in Tables 20-22, for the adult age span (i.e., 16-69 years) in Tables 23-25, and for the older age span (i.e., 60-89 years) in Tables 26-28. Each set of tables includes the multivariate base rates for the full D-KEFS battery and the two brief batteries, with stratifications by estimated level of intelligence and years of education.

When using multivariate base rates to understand a test profile, the following guidelines can be helpful for the interpretation of the cumulative percentages of low scores on the D-KEFS: *common* (≥25% of the normative sample), *fairly uncommon* (10% to 24% of normative sample), *uncommon* (5% to 9% of the normative sample), and *very uncommon* (<5% of the normative sample). As an example, 91.9% of participants within the total sample (i.e., ages 16-89) had one or more scores at or below the highest cutoff evaluated (i.e., ≤25th percentile). Using this same cutoff, 28% of the total had six or more low scores. In turn, for individuals aged 16 to 89 completing the entire D-KEFS battery (i.e., 16 primary test scores), it is *common* to have 0-6 scores at or below the 25th percentile. Even when using the lowest cutoff score (i.e., ≤2nd percentile), 36.9% of the normative sample had one or more scores at or below this cutoff when administered the full D-KEFS battery, indicating that a score labeled as impaired when interpreted in isolation is actually quite *common* among adult participants administered the full D-KEFS battery. However, it is *fairly uncommon* to have two scores at or below the 2nd percentile (cumulative percentage = 16.6%), *uncommon* to have three scores at or below the 2nd percentile (cumulative percentage = 7.7%), and *very uncommon* to have four or more low scores at or below the 2nd percentile (cumulative percentage = 4.1%).

As seen in the tables, low scores occurred quite commonly among the normative sample of the D-KEFS. For the total sample, 82.6% obtained one or more low scores at or below the 16th percentile when administered the full test battery (see Table 20). The prevalence of low scores decreased in tandem with the level of the cutoff score. Again for the total sample administered the full D-KEFS battery, 50.7% obtained one or more low scores at or below the 5th percentile and 36.9% obtained one or more low scores at or below the 2nd percentile (see Table 20). As anticipated, these base rates were lower for the brief batteries than for the full battery. When interpreting the Total Achievement scores for the four-test and three-test batteries, 71.8% and 62.8% of the total sample, respectively, obtained at least one score at or below the 16th percentile (see Tables 21 and 22, respectively). Even with the lowest cutoff (i.e., ≤2nd percentile), 26.8% of the sample obtained at least one or more low scores for the three-test battery (see Table 22), and 30.1% of the sample obtained at least one or more low scores for the four-test battery (see Table 21). As made clear by this example, scores in the univariate impaired range occur quite commonly when interpreting the D-KEFS scores in isolation, even when evaluating scores from just a short battery of tests.

In contrast to the number of tests interpreted, the age of participants had little influence on the prevalence of low scores. When comparing the adult (i.e., 16-69 years) and older adult samples (i.e., 60-89 years), there was no discernible differences in prevalence of low scores. Using the four-test battery (i.e., 9 scores), 71.4% of the sample aged 16-69 (see Table 24) and 73.6% of the sample aged 60-89 (see Table 27) had one or more scores at or below the 16th percentile. Again for the four-test battery, among the 16-69 and 60-89 year age groups, 44.8% and 43.8%, respectively, had two or more scores at or below the 16th percentile (see Tables 24 an 27, respectively). This pattern held for a more extreme low score cutoff as well. At the 2nd percentile cutoff, it was *common* in both age spans to have one or more low scores and *very uncommon* to have three or more low scores.

Aside from age, other stratifications had their expected relationships with the prevalence of low scores. The estimated level of intelligence from the WASI was related to the base rates of low scores on the D-KEFS, where the number of low scores decreased with increased WASI FSIQ (e.g., see Tables 20-28). Figure 11 shows this relationship visually for the three-test D-KEFS battery (i.e., 7 scores) among the total sample. For this same battery, Figure 12 shows that years of education followed a similar inverse relationship with low score frequencies, where individuals with fewer years of education had more low scores than those with more years of education.

Discussion

As with other neuropsychological test batteries (Binder et al., 2009; Brooks et al., 2010, 2012, 2013a), and as expected, low scores occur commonly among individuals administered the D-KEFS. The frequency of low scores on the D-KEFS was related to the cutoffs used to define a low score and the number of scores interpreted; however, even when using the lowest score cutoff (i.e., ≤2nd percentile) and considering the briefest battery evaluated (i.e., three-test battery), obtaining at least one low score still occurred quite *commonly* among individuals in the total normative sample (i.e., cumulative percentage = 26.8%). In comparison to other frequently administered test batteries, low scores occurred more often among the subtests of the D-KEFS when the entire battery is administered. For example, using a typical low score cutoff of less than or equal to the 16^{th} percentile, 82.6% of the total sample had at least one low score when administered the full D-KEFS; however, using this same cutoff, only 62.8% and 63.8% of the WAIS-IV and WMS-IV normative samples, respectively, had one or more low scores (Brooks et al., 2013a). This is expected, of course, because the full D-KEFS has 9 tests and provides 16 primary scores, while the WAIS-IV and WMS-IV each consist of 10 tests with 10 primary scores.

When considering differences in low score frequencies across batteries, the number of tests administered and interpreted should be considered. When interpreting fewer D-KEFS scores, the low score frequencies are more similar to WAIS-IV and WMS-IV base rates. For the four-test D-KEFS battery, 71.8% of the adult normative sample had one or more low scores; and for the three-test D-KEFS battery, 62.8% had one or more low scores (i.e., both using the $\leq 16^{th}$ percentile cutoff). However, when using a lower percentile cutoff, the prevalence of low scores is greater for the D-KEFS than the WAIS-IV and WMS-IV, even when fewer scores are interpreted. Consider the lowest cutoff evaluated (i.e., $\leq 2^{nd}$ percentile) in the following example. When interpreting the ten primary subtests on the WAIS-IV, 12.6% of the normative sample (i.e., ages 16-90) had one or more low scores (Brooks et al., 2013a). When interpreting the ten subtests on the WMS-IV, 21% of the adult normative sample (i.e., ages 16-69) had one or more low scores (Brooks et al., 2013a). In contrast, for the total D-KEFS sample (i.e., ages 16-89), 30.1% had one or more low scores on the four-test battery (i.e., 9 test scores). Figure 13 displays a comparison between the WAIS-IV, WMS-IV, and D-KEFS in terms of low score prevalence for multiple low score cutoffs. The figure shows the frequency of one or more low scores *divided by the number of tests interpreted*; demonstrating that, per test interpreted, the brief D-KEFS batteries had a greater frequency of low scores than both the WAIS-IV and WMS-IV. Figure 14 compares the base rates of low scores for the WAIS-IV (10 scores), WMS-IV (10 scores), and the four-test D-KEFS battery (9 scores). As displayed visually, the four-test D-KEFS battery resulted in higher base rates of low scores, even when interpreting one less score than the WAIS-IV and WMS-IV batteries.

The greater rate of low scores on the D-KEFS battery may derive from both the nature of executive functions and/or the psychometrics of executive function assessment. The D-KEFS has received criticism in the past due to the low reliability estimates reported in the technical manual for many of its tests (Schmidt, 2003), although the reliability of the D-KEFS is not uncommonly low for executive function tests or neuropsychological measures in general (Delis, Kramer, Kaplan, & Holdnack, 2004). When people are administered multiple D-KEFS measures, fluctuations in scores could derive in part from measurement error, where the D-KEFS tests are tapping into abilities unrelated to executive-related constructs. Some of this error may be due to the metric of the D-KEFS test scores. Notably, for the three-test and four-test batteries, the three Total Achievement scores from the Trail Making Test and Color-Word Interference Test are time-based, scored in seconds to completion; and fluctuations in these performances may come from distractions or variable effort and engagement during administration rather than true executive-related deficits.

As opposed to measurement error impacting performance variability, executive function could also be more variable due to the nature of the construct itself. A rich history of research has explored the overlap between executive functions and intra-individual variability. Notably, executive functions and cognitive variability are both related to the efficiency of frontal lobe functioning (MacDonald, Nyberg, & Bäckman, 2006), indicating that clinicians may observe more variable performances among tests evaluating higher-order cognitive abilities. As a consequence, typical variability in executive functions could explain the greater frequency of scores falling in the univariate impaired range for the D-KEFS normative sample. Normal fluctuations in executive functions may appear as poor individual performances across some measures over the administration of an extensive test battery. In a typical comprehensive neuropsychological assessment, clinicians could likely administer the WAIS-IV, WMS-IV, and a brief D-KEFS battery, which would result in an extremely high likelihood of finding one or more impaired scores among their results.

With a high likelihood of low scores occurring when healthy adults are administered a battery of tests, neuropsychologists must be wary of over-interpreting impaired scores in isolation (Brooks et al., 2013a). For instance, past researchers have focused on the misdiagnosis of amnestic Mild Cognitive Impairment (MCI) when administering multiple memory tests to older adults, finding that low scores on memory tests occur quite commonly when interpreting a full battery of memory test scores (Brooks et al., 2007, 2008). The diagnostic criteria for MCI include (a) self- or informant-reported cognitive complaints, (b) objective cognitive impairment, (c) intact functional independence, and (d) the absence of a dementia diagnosis (Albert et al., 2011; Peterson, 2004; Winblad et al., 2004). When defining objective cognitive impairment, researchers and clinicians have begun to consider impaired performances on tests evaluating multiple cognitive domains in addition to memory, such as executive functions. This conceptual change in the domains considered in the evaluation of MCI has resulted in three additional diagnostic constructs: amnestic multi-domain MCI, non-amnestic multi-domain MCI, and non-amnestic single-domain MCI (Peterson, 2004; Peterson et al., 2014; Winblad et al., 2004).

When paired with other relevant client information (e.g., medical and psychiatric history, cognitive complaints), these three MCI subtypes could be diagnosed based on one or more impaired performances on a test of executive functions (Petersen et al., 2014); and, based on the current findings, low scores occur fairly often in the context of a multi-test assessment of executive functions. For example, during a typical assessment that includes the evaluation of executive functions, some of the most commonly administered tests include the Trail Making Test, Verbal Fluency Test, Color Word Interference Test, and Tower Test (Rabin et al., 2016). When interpreting the nine D-KEFS Total Achievement scores for just these four tests, the likelihood of identifying a score in the impaired range is higher with the D-KEFS than with other test batteries (see Figures 13 and 14). In turn, the false-positive rate for an MCI diagnosis is potentially higher when evaluating older adult clients for executive dysfunction as opposed to impairment in other areas of cognition.

With the MCI scenario as an example, in order to prevent the misdiagnosis of MCI, clinicians can consider the frequency of low scores on the D-KEFS in combination with multivariate base rates for other commonly administered batteries. Together, the WAIS-IV, WMS-IV, and D-KEFS are the most commonly administered test batteries in neuropsychological practice (Rabin et al., 2016), and clinicians now have access to published multivariate base rates for all three batteries. As with the base rate derivations for the WAIS-IV and WMS-IV (Brooks et al., 2013a), both education and intelligence were related to low score frequencies, where the prevalence of low scores decreased with higher levels of intelligence and more years of education. Combining advanced psychometric and demographic information, researchers can now base their multivariate interpretation of test data on a comprehensive set of quantitative resources.

In the context of an MCI assessment, consider a 70-year-old client with a high school education completing the ten primary subtests of the WAIS-IV, six subtests of the WMS-IV Older Adult battery (i.e., Logical Memory I & II, Verbal Paired Associates I & II, and Visual Reproduction I & II), and three subtests of the D-KEFS (i.e., the Trail Making, Verbal Fluency, and Color Word Interference Tests). Using stratifications for years of education and a common low score cutoff (i.e., ≤16th percentile), 70.7% had one or more low scores on the WAIS-IV (see Table eA2.14 in Brooks et al., 2013a), 53.4% had one or more low scores on the WMS-IV (see Table eA2.20 in Brooks et al., 2013a), and 65.4% had one or more low scores on the D-KEFS (see Table 28 in the current chapter). Using the same normative tables at a lower score cutoff (i.e., ≤5th percentile), 30.9% had one or more low scores on the WAIS-IV, 23.9% had one or more low scores on the WMS-IV, and 40.2% had one or more low scores on the D-KEFS.

Under the diagnostic criteria for MCI, objective cognitive impairment "is defined as a poor performance in one or more cognitive measures, which suggests deficits in one or more cognitive areas or domains" (Petersen et al., 2014, p. 217). When administered multiple batteries, the potential for a healthy adult meeting the MCI criterion for objective cognitive impairment appears extraordinarily high when using a fairly lenient low score cutoff, and still fairly high when considering a lower cutoff that is commonly used in the assessment of MCI (i.e., $\leq 5^{th}$ percentile). Based on the criteria described by Petersen and colleagues, roughly two in five healthy older adults would produce at least one score in the univariate impaired range; and without context regarding the normal prevalence of low scores, a clinician could misinterpret these findings as suggesting true executive function deficits, meeting the psychometric criterion for MCI based on D-KEFS performances alone.

The MCI example illustrates the usefulness of multivariate base rates when interpreting D-KEFS test data. However, just as an MCI diagnosis cannot be based on psychometric scores alone, the base rates should only be interpreted alongside a client's psychological and medical history. As well, a recent simulated re-analysis of WMS-IV normative data demonstrated that an appreciation for the assessment context (i.e., most clients that present in a clinical setting tend to have cognitive impairment; Lonie et al., 2008) and the frequency of cognitive impairment among clinical samples could improve a clinician's accuracy at identifying true cognitive impairment (Gavett, 2015). Unfortunately, these sources of information are not easily available. Each clinic likely has a different frequency of cognitive impairment, and the D-KEFS normative data did not include a mixed clinical sample from which multivariate base rates can be derived. Nevertheless, an appreciation for these factors in future studies could improve the clinical utility of multivariate base rate research.

The base rates presented herein can be useful in any context where multiple D-KEFS tests are administered and interpreted among adult clients; and future publications will focus specifically on pediatric assessment, evaluating low score frequencies among the pediatric portion of the normative sample not considered in the current results. Executive dysfunction has been implicated as a key element in many neuropsychological disorders often diagnosed in childhood, including ADHD (Brown, 2008), autism, (Kenworthy, Yerys, Anthony, & Wallace, 2008), and Fetal Alcohol Spectrum Disorder (FASD; Khoury, Milligan, & Girard, 2015). Considering the prevalence of low D-KEFS scores among adults, clinicians may also be at-risk of over-interpreting low D-KEFS scores in pediatric assessments, which could result in the misdiagnosis of neurodevelopmental conditions that consider executive functions in diagnostic decision-making (e.g., FASD; Chudley et al., 2005; Hoyme et al., 2005). Past multivariate research has evaluated low score frequencies among pediatric populations (Brooks & Iverson, 2012), focusing on tests designed to specifically evaluate children and adolescents (e.g., the Wechsler Intelligence Scale for Children, Fourth Edition, Brooks, 2010; the NEPSY, Brooks et al., 2010). Pediatric base rates of low score frequencies for the D-KEFS would contribute to this research, and provide some of the first multivariate base rates for an executive function test battery with pediatric norms.

Epilogue

Entitled *Towards a Multivariate Assessment of Executive Functions*, all three chapters of this dissertation shared a common theme of approaching executive functions through a multivariate framework, attempting to summarize measurement approaches and translate these approaches into clinical practice in a meaningful way. As stated in the concluding paragraph of Chapter 1, the findings of the first systematic review and re-analysis served as an affirmation of the elusive nature of this construct; and – in many ways – this conclusion applies to the dissertation in its entirety. The first chapter showed some modest support for a three-factor or nested factor model among adults and a non-differentiated shifting factor early in life, but it also elucidated potential publication biases and replicability concerns for this research field. The second chapter attempted to link research and clinical practices using a common assessment battery (i.e., the D-KEFS), but the use of residualized scores precluded the direct translation of the findings into clinically applicable outcomes. Lastly, the third chapter resulted in a direct clinical application through the derivation of multivariate base rates for the D-KEFS, but the calculation of these base rates did not involve a theory-based approach.

As noted in the Prologue, the impetus for this dissertation was the disconnect I observed between research and clinical practice in regards to how neuropsychologists evaluated executive functions. Clinical neuropsychology stands as an applied science; and as a scientist-practitioner, I aimed to bridge science and practice as best I could in the final steps before entering my career. However, the conclusions of this dissertation are but a first step in a long scientific journey *towards* the multivariate assessment of executive functions – a journey that will improve the psychometric evaluation of this elusive construct in both research and clinical settings, and ultimately enhance our understanding of the self-regulatory capacity of the human mind.

Cited Figures

Figure 1. Flowchart of systematic review

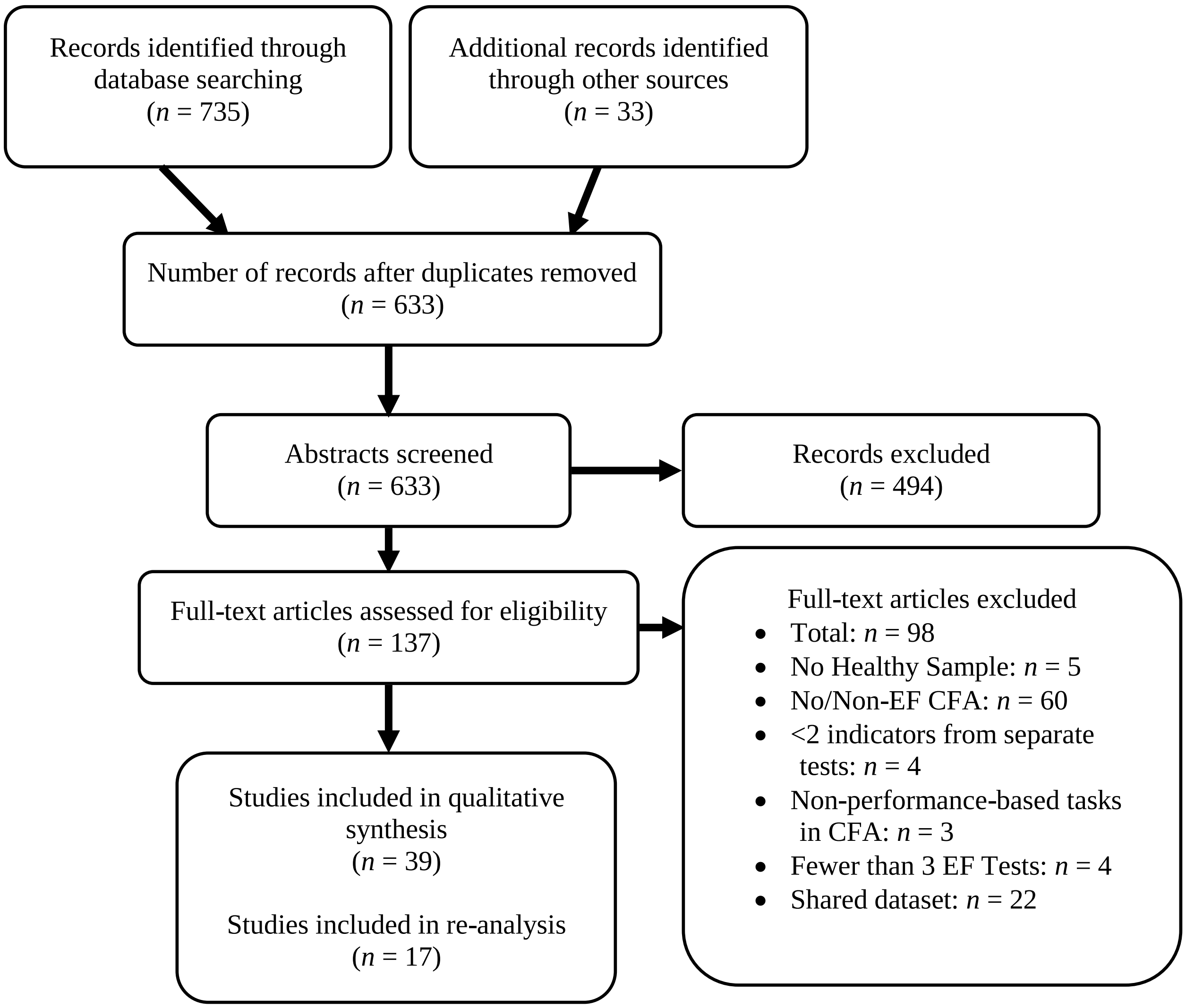


Figure 2. Diagrams of factor models tested in the re-analysis

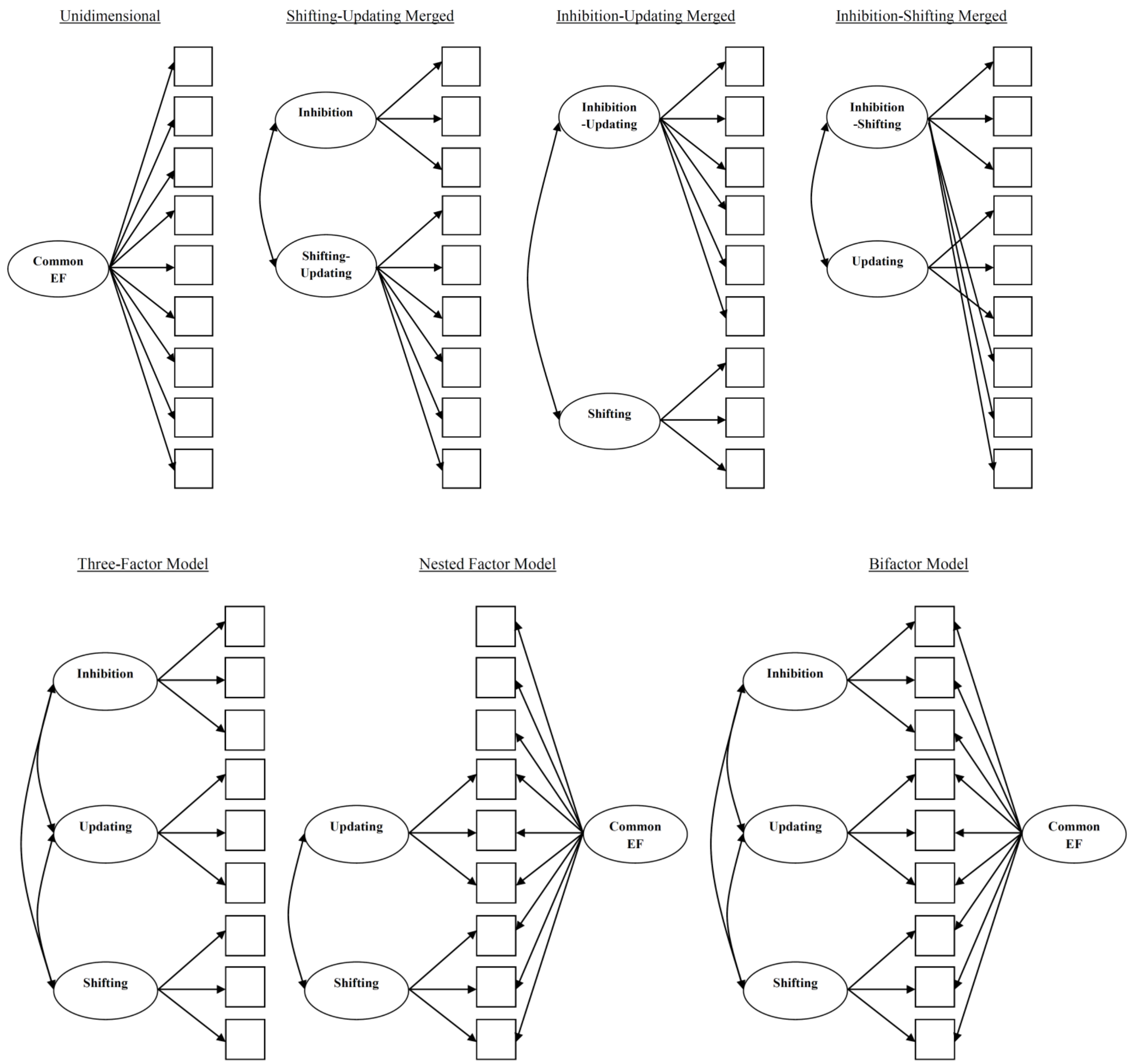


Figure 3. Child and Adolescent Studies: Forest Plot of Average Percent Convergence among 5,000 Bootstrapped Samples by Measurement Model

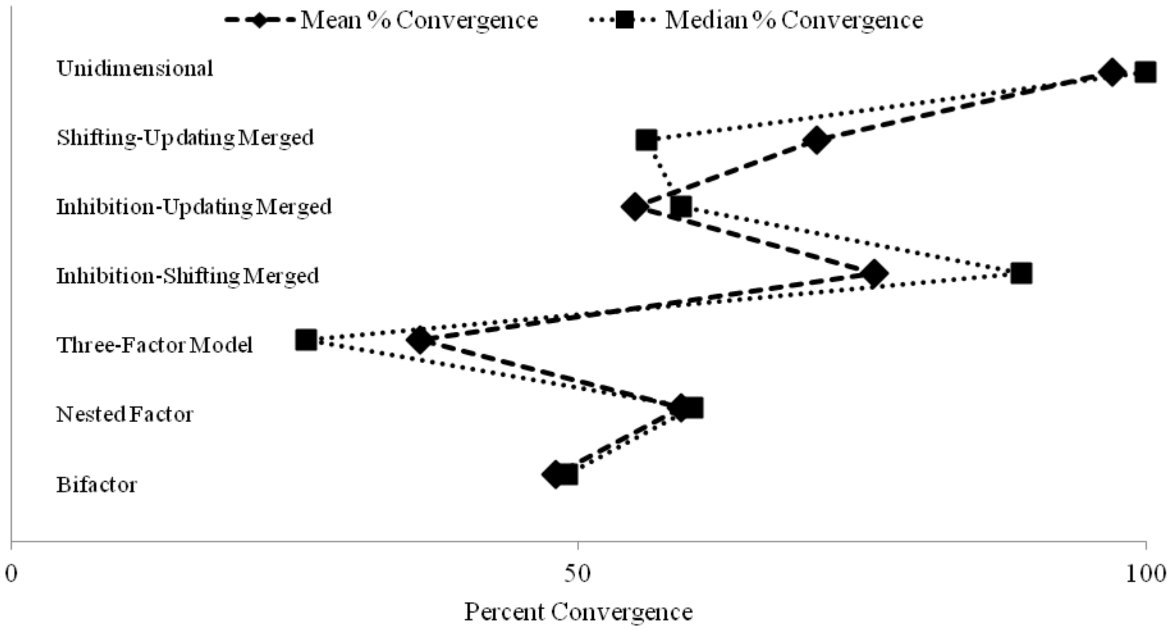


Figure 4. Adult Studies: Forest Plot of Average and Median Percent Convergence among 5,000 Bootstrapped Samples by Measurement Model

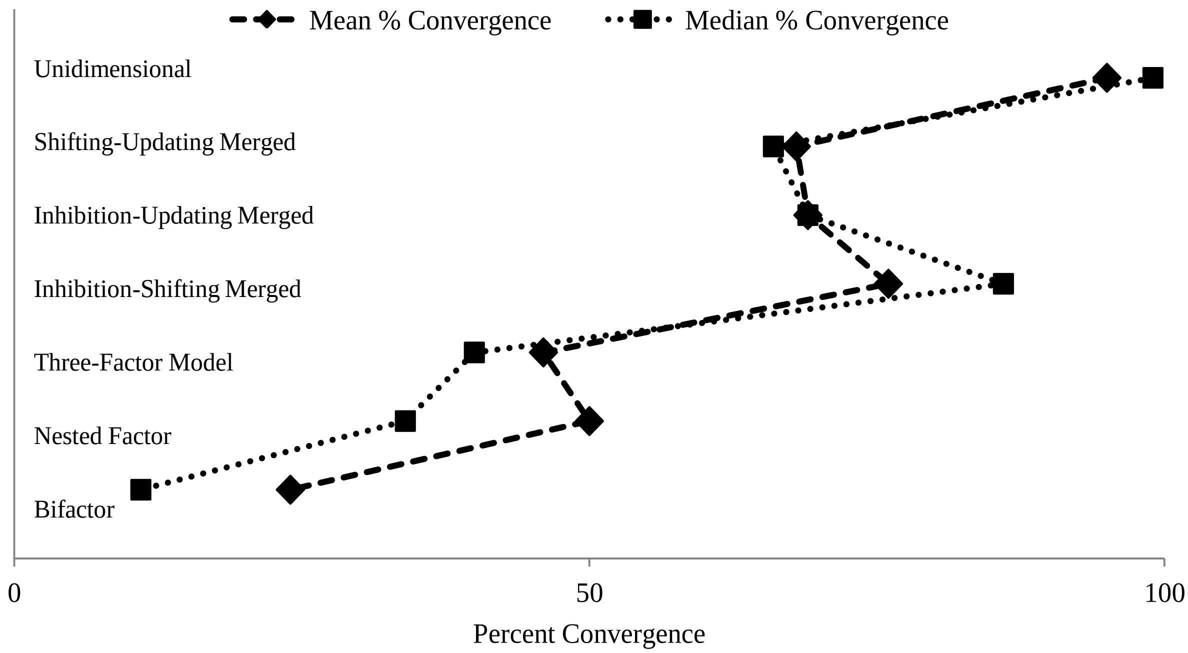
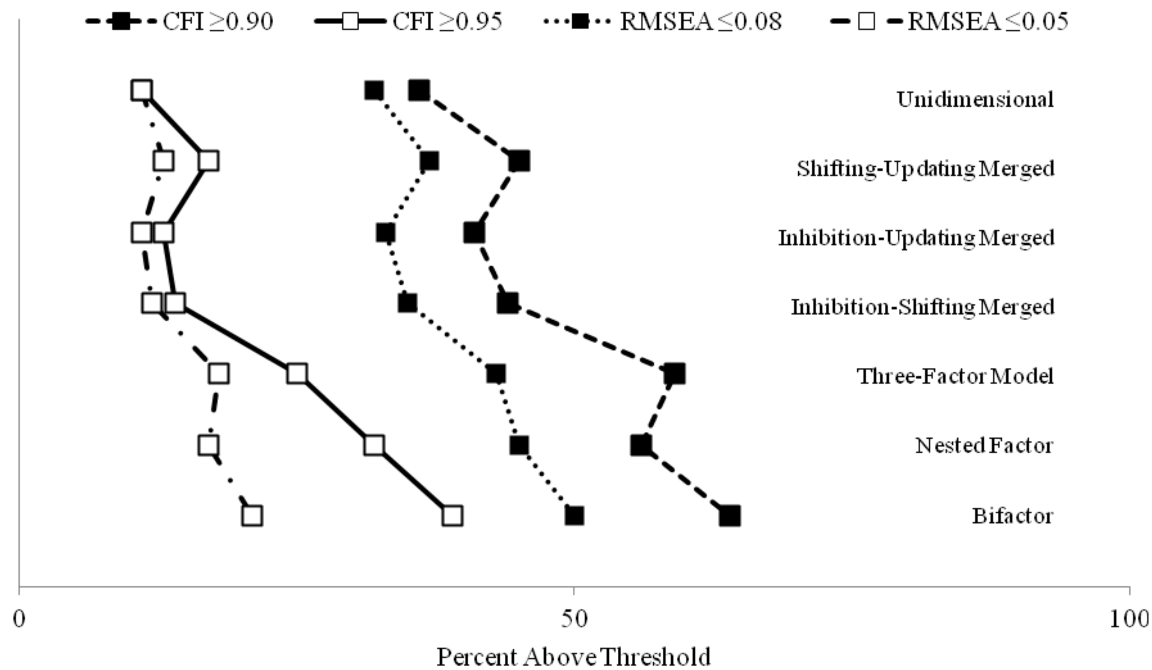
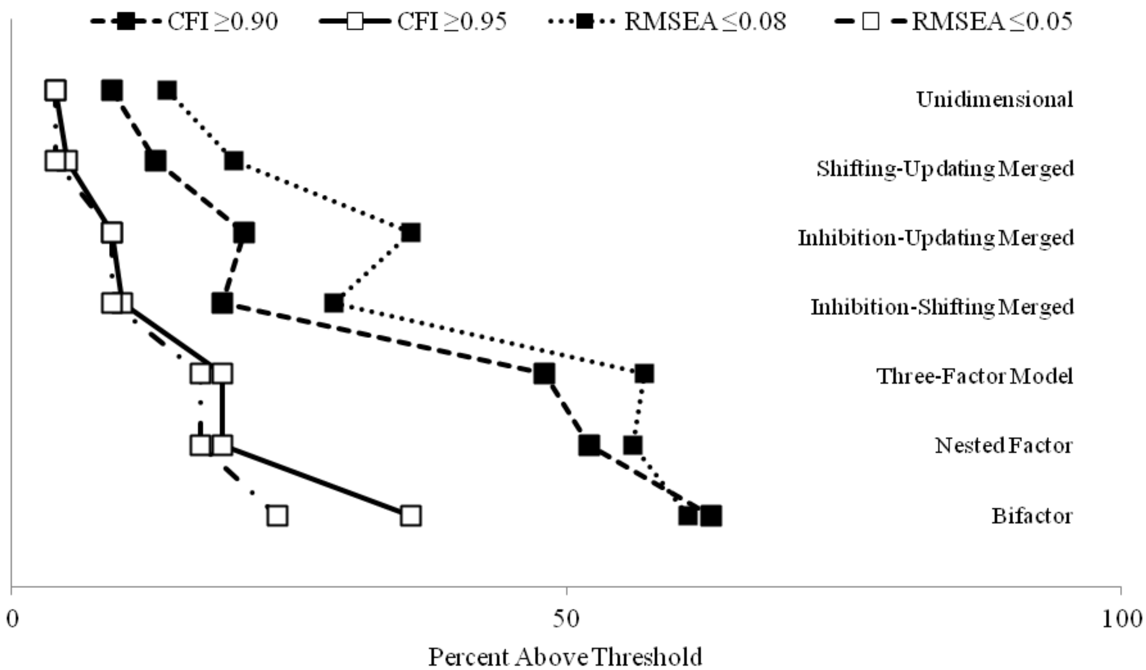


Figure 5. Child and Adolescent Studies: Forest Plot of Average Percent of Converged Models Meeting Lenient (i.e., CFI ≥ 0.90 and RMSEA ≤ 0.08) or Strict (i.e., CFI ≥ 0.95 and RMSEA ≤ 0.05) Fit Criteria by Measurement Model



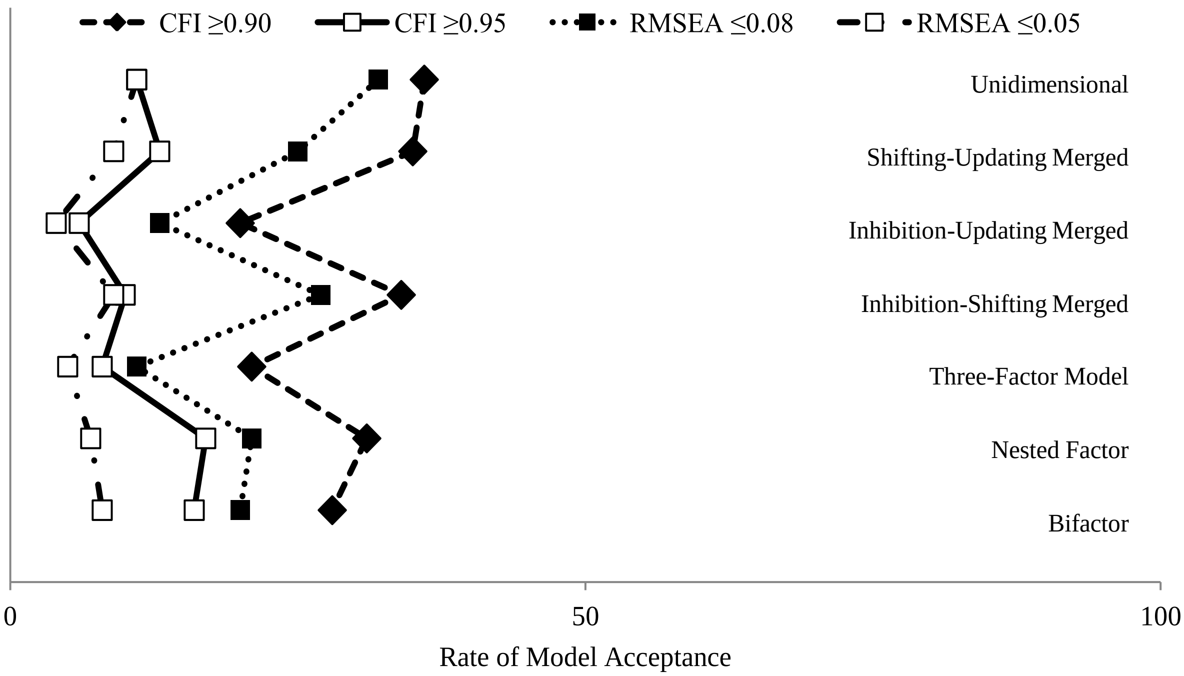
Note. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

Figure 6. Adult Studies: Forest Plot of Average Percent of Converged Models Meeting Lenient (i.e., CFI ≥ 0.90 and RMSEA ≤ 0.08) or Strict (i.e., CFI ≥ 0.95 and RMSEA ≤ 0.05) Fit Criteria by Measurement Model



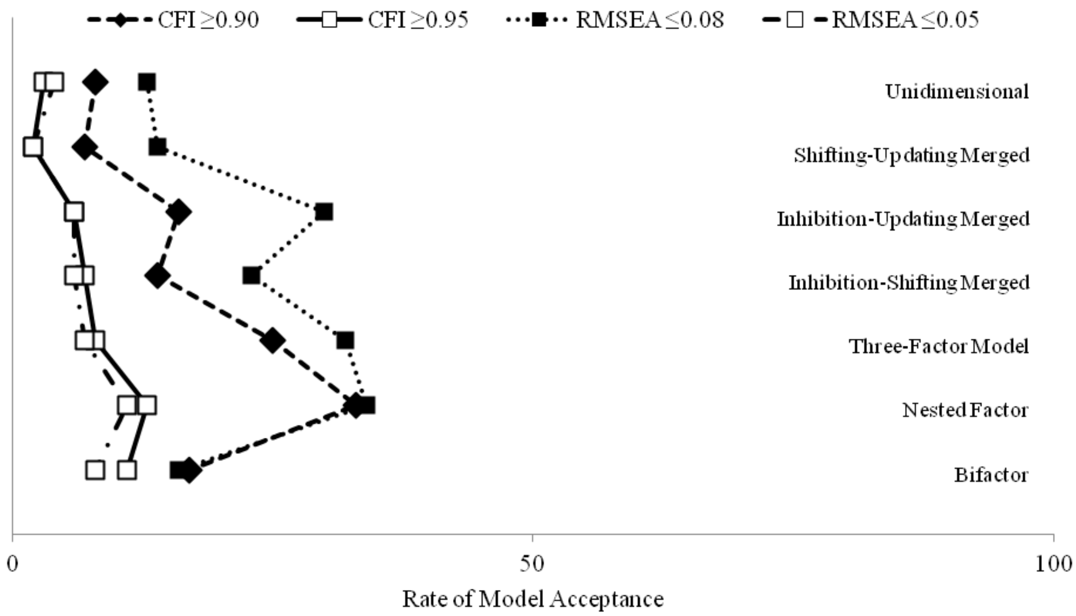
Note. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

Figure 7. Child and Adolescent Studies: Forest Plot of Percent of Models Both Converging and Meeting Lenient (i.e., CFI ≥ 0.90 and RMSEA ≤ 0.08) or Strict Fit Criteria (i.e., CFI ≥ 0.95 and RMSEA ≤ 0.05) by Measurement Model



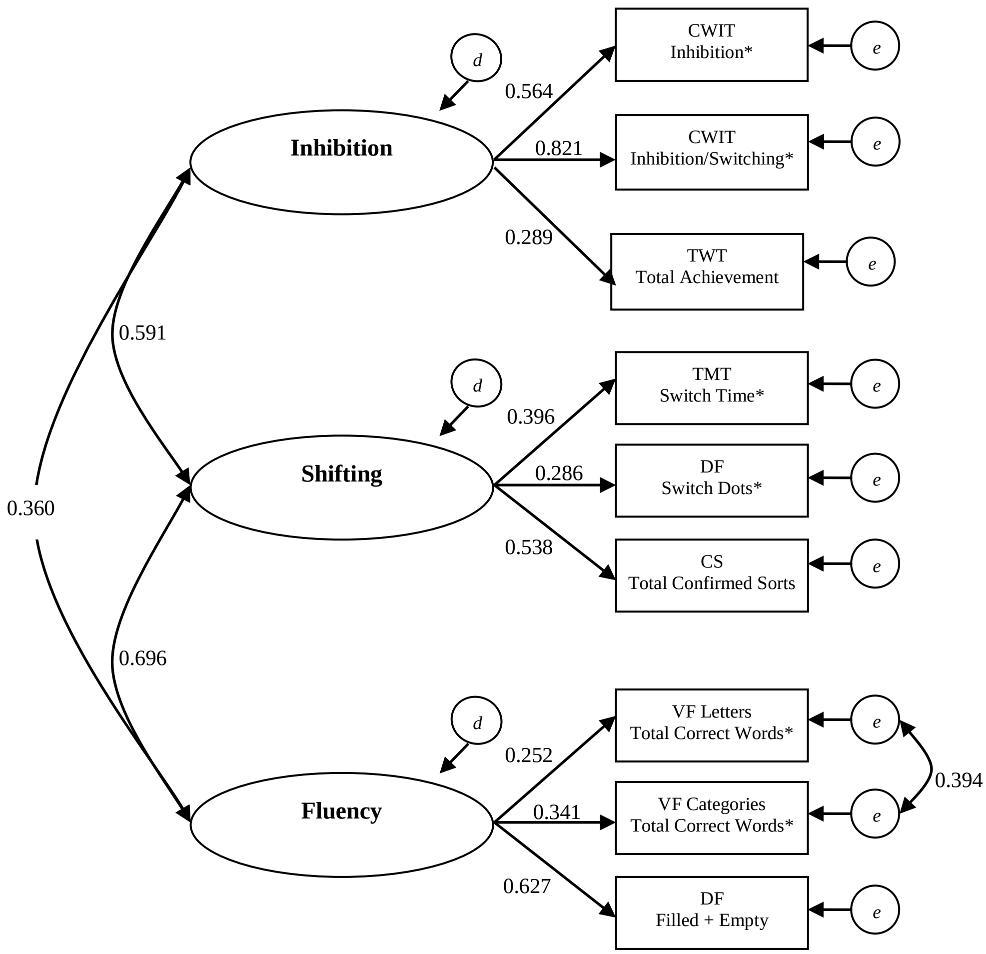
Note. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

Figure 8. Adult Studies: Forest Plot of Percent of Models Both Converging and Meeting Lenient (i.e., CFI ≥ 0.90 and RMSEA ≤ 0.08) or Strict Fit Criteria (i.e., CFI ≥ 0.95 and RMSEA ≤ 0.05) by Measurement Model



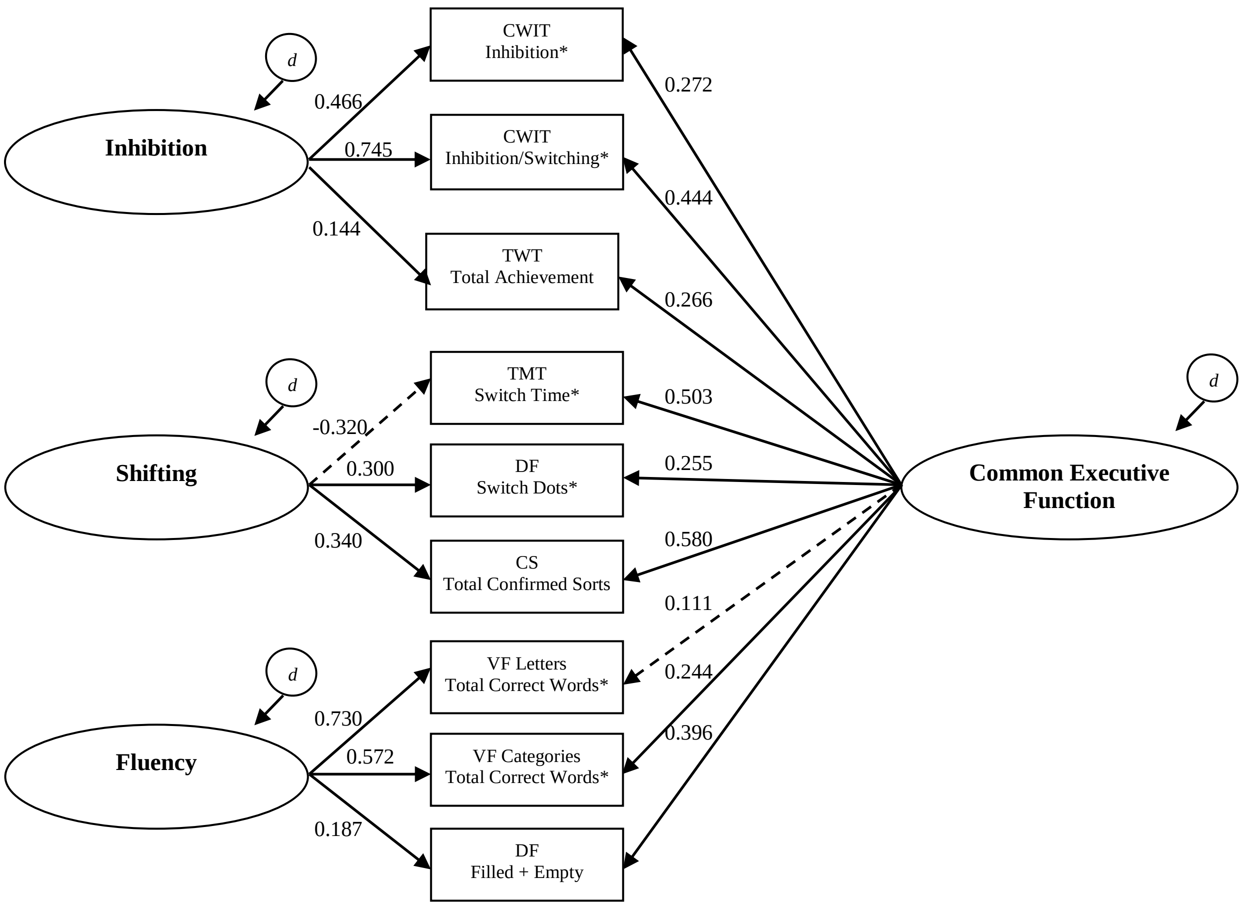
Note. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

Figure 9. Three-Factor First-Order Measurement Model



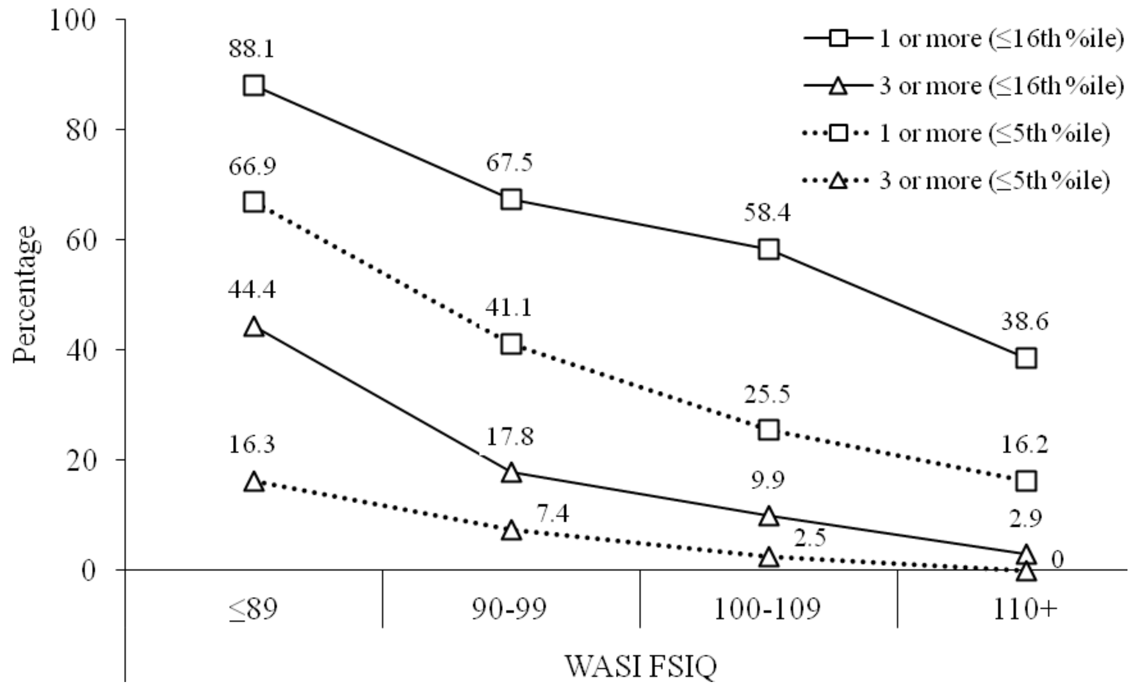
Note. CS = Card Sorting Test; CWIT = Color-Word Interference Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency. *Indicates a value that was residualized of variance attributable to a control variable. Errors are not displayed above, but were as follows: CWIT Inhibition = 0.682; CWIT Inhibition/Switching = 0.326; TWT Total Achievement = 0.917; TMT Switch Time = 0.843; DF Switch Dots = 0.918; CS Total Confirmed Sorts = 0.710; VF Letters Total Correct Words = 0.936; VF Categories Total Correct Words = 0.884; DF Filled + Empty = 0.607

Figure 10. Bifactor Factor Measurement Model.



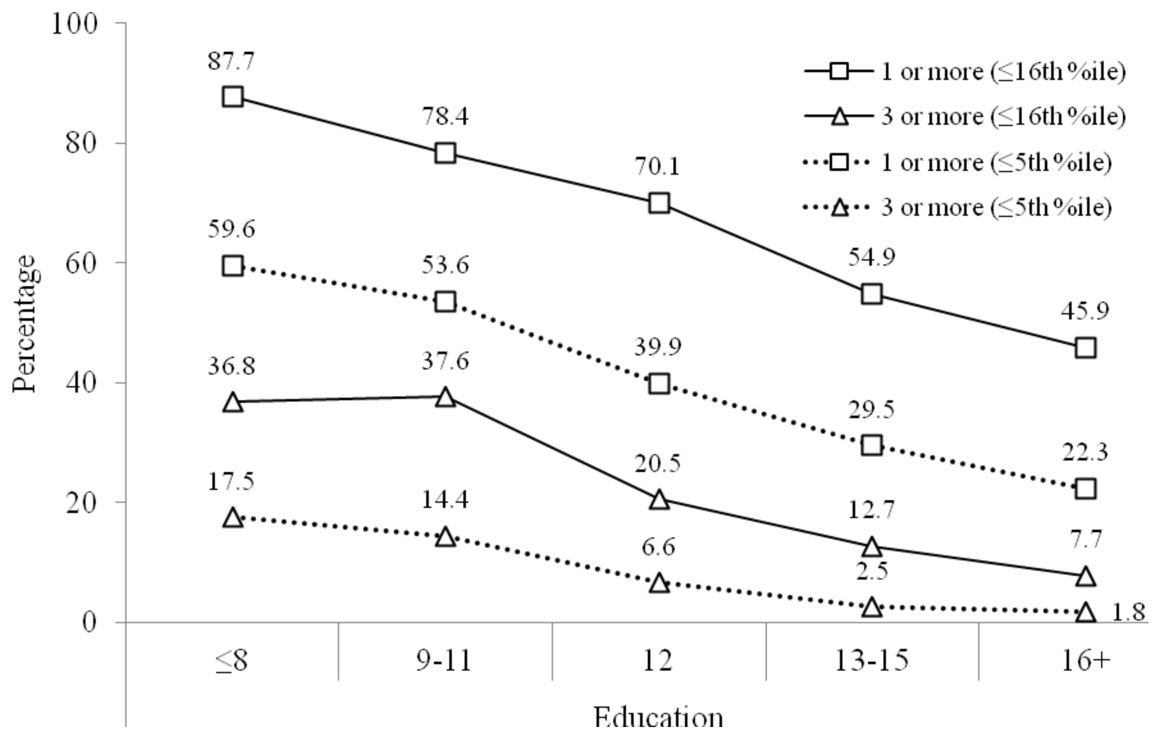
Note. CS = Card Sorting Test; CWIT = Color-Word Interference Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency. *Indicates a value that was residualized of variance attributable to a control variable. Dashed lines designate non-significant loadings. Errors are neither displayed nor diagrammed above, but were as follows: CWIT Inhibition = 0.709; CWIT Inhibition/Switching = 0.248; TWT Total Achievement = 0.909; TMT Switch Time = 0.644; DF Switch Dots = 0.845; CS Total Confirmed Sorts = 0.547; VF Letters Total Correct Words = 0.455; VF Categories Total Correct Words = 0.614; DF Filled + Empty = 0.809

Figure 11. The frequency of low D-KEFS scores is contingent on level of estimated intelligence (cut-offs: ≤16th percentile and ≤5th percentile)



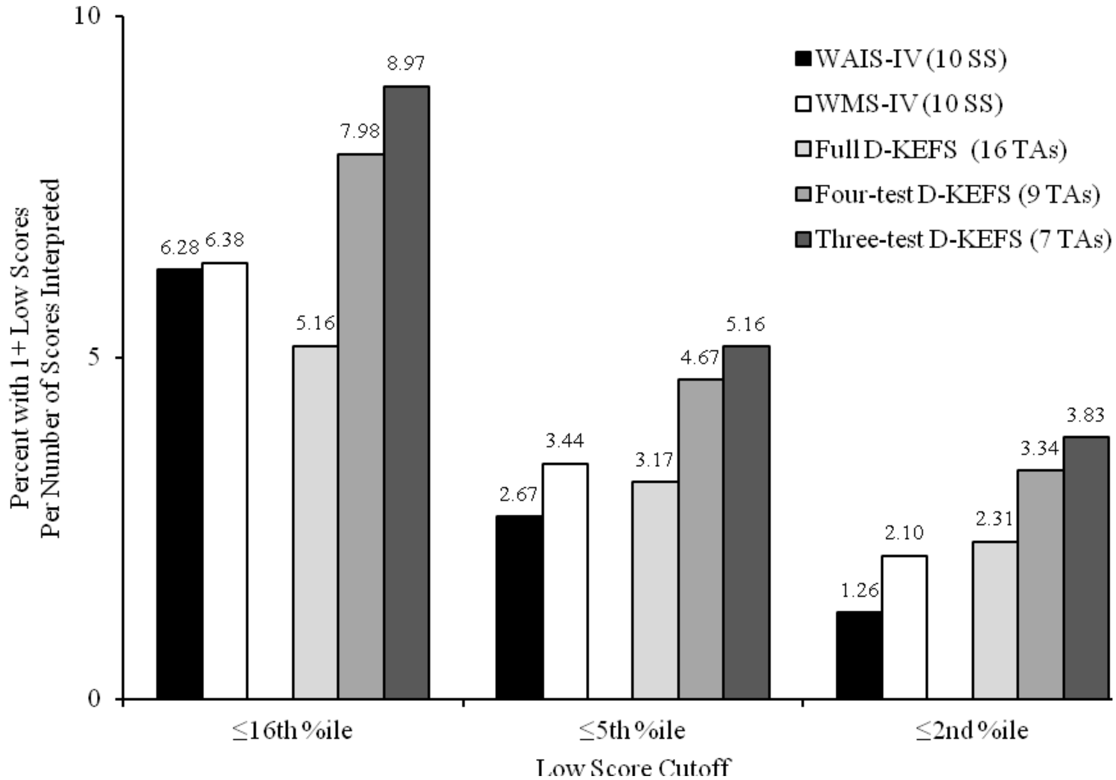
Note. The data included in this figure was based on interpretation of the three-test D-KEFS battery (7 Total Achievement Scores) for the sub-sample of participants aged 16-89 years old administered the WASI FSIQ.

Figure 12. The frequency of low D-KEFS scores is related to years of education (cut-offs: ≤16th percentile and ≤5th percentile)



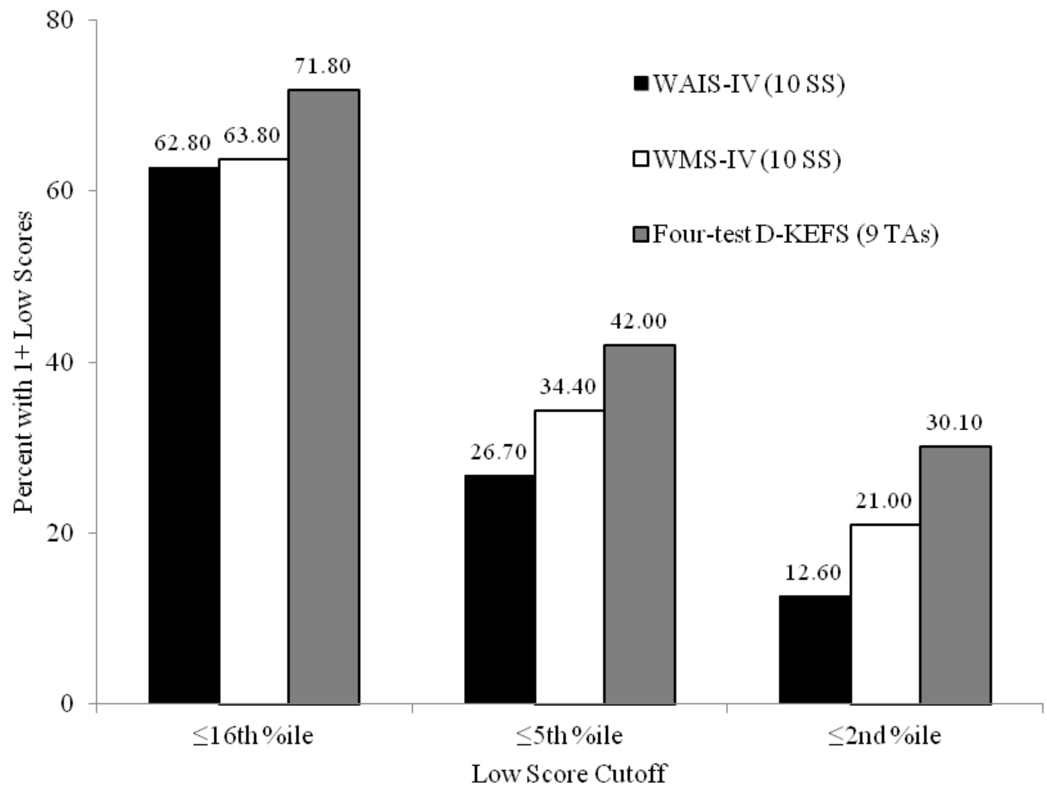
Note. The data included in this figure was based on interpretation of the three-test D-KEFS battery (7 Total Achievement Scores) for the participants aged 16-89 years.

Figure 13. Percentage of the normative samples for the WAIS-IV (ages 16-90), WMS-IV (ages 16-69), and D-KEFS (ages 16-89) with one or more low scores, based on three cutoffs (i.e., ≤16th ≤5th, and ≤2nd percentile), per number of test scores interpreted.



Note. The WAIS-IV and WMS-IV values included in this figure were calculated based on data derived from another publication on multivariate base rates (Brooks et al., 2013a).

Figure 14. Percentage of the normative samples for the WAIS-IV (ages 16-90, 10 scores), WMS-IV (ages 16-69, 10 scores), and D-KEFS Four-Test battery (ages 16-89, 9 scores) with one or more low scores, based on three cutoffs (i.e., ≤16th ≤5th, and ≤2nd percentile)



Note. The WAIS-IV and WMS-IV values included in this figure were obtained from another publication on multivariate base rates (Brooks et al., 2013a).

Cited Tables

Table 1. *Studies Reporting Measurement Models of Executive Functions: Sample Characteristics and Study Quality*

Age Group ($\bar{x}$ age)	Author	*N*	Age (years): $\bar{x}$ (*SD*)	Age Range (years)	% Female	% White or Category	Education: $\bar{x}$ (years) or Category	Study Quality*
Preschool (<6 yrs.)	Carlson et al. (2014)	104	4.00 (0.43)	3-5	46.15	80.00	Some preschool	9
	Lerner & Lonigan (2014)	289	4.65 (0.63)	3-5	53.00	57.00		8
	Masten et al. (2012)	138	5.77 (0.58)	4.83-6.92	56.50	4.30		6
	Miller et al. (2012)[+]	129	4.17 (0.58)	3-5	39.53	80.00		9
	Monette et al. (2015)	272	5.70 (0.34)		54.55			10
	Usai et al. (2014)[+]	175	5.71 (0.28)	5-6	43.43		Kindergarteners	9
	Wiebe et al. (2008)	243	3.92 (1.00)	2-6	55.56	70.37	Preschool	9
	Wiebe et al. (2011)	228	3.01 (0.04)		49.56	75.88		9
	Willoughby et al. (2012a)	1036	5.03 (0.26)		50.00			8
School-Aged (6-12 yrs.)	Agostino et al. (2010)[+]	155	10.08 (1.25)	8-13	56.00		21.94% Grade 3; 25.81% Grade 4; 28.39% Grade 5; 23.87% Grade 6	7
	Arán-Filippetti (2013)[1+]	LSES: 124 MSES: 124		8-12			Some grade school	8
	Brocki & Tillman (2014)	114	9.32 (2.31)	5-14	47.00	68.00		10
	Brydges et al. (2012)[+]	215	8.33 (1.08)	7-9	48.84			9
	Duan et al. (2010)[+]	61	11.88 (0.65)	11-12	44.27	From “Chinese schools”		7
	Lambek & Shevlin (2011)[2]	164		7-12	47.00			5
	Lee et al. (2012)	163	6.90 (0.31)		49.69	“Mainly ethnic Chinese”		8
	Lee et al. (2013)[3]	332		All 8				9
	Lehto et al. (2003)[+]	103	10.50 (1.30)	8-13	44.44			8
	Rose et al. (2012): Full-Term[4+]	131	11.14 (0.35)	10.5-12.5	49.30	16.10		8
	Rose et al. (2012): Pre-Term[4+]		11.18 (0.44)	10.4-12.1		11.40		

Age Group ($\bar{x}$ age)	Author	*N*	Age (years): $\bar{x}$ (*SD*)	Age Range (years)	% Female	% White or Category	Education: $\bar{x}$ (years) or Category	Study Quality[*]
	van der Sluis et al. (2007)	172	10.67 (0.72)	9-12	51.16		58.14% - Grade 4; 41.86% - Grade 5	10
	van der Ven et al. (2013)	211	6.42 (0.37)	5-7	47.87			11
	Xu et al. (2013)[5+]	140	8.78 (0.57)	7-9	47.86	"Chinese"		8
		165	11.59 (0.88)	10-12	49.09	"Chinese"		
Adolescents (13-17 yrs.)	Friedman et al. (2011) [+]	813	17.30 (0.60)	15-20	51.66			11
	Lambek & Shevlin (2011)[2]	75		13-16	45.00			5
	Xu et al. (2013)[5+]	152	14.41 (0.86)	13-15	50.00	"Chinese"		8
Adults (18-59 yrs.)	Chuderski et al. (2012)	160	21.90 (2.70)	15-35	61.25			8
	Fleming et al. (2016)[+]	420	22.50		50.00	91.00		10
	Fournier-Vicente et al. (2008)	180	23.62 (3.24)	18-31	50.00		Undergraduates	11
	Ito et al. (2015)[+]	484	19.75 (2.21)	18-42	49.18	86.16	Undergraduates	11
	Klauer et al. (2010)[6] - Study 1[+]	125	23.10 (5.80)	17-57	42.97			8
	Klauer et al. (2010)[6] - Study 2[+]	118	23.50 (4.20)	18-42	35.25			9
	Miyake et al. (2000)[+]	137					Undergraduates	9
	Was (2007)	188	25.70	18-56	70.21		Undergraduates	8
Older Adults (>60 yrs.)	Adrover-Roig et al. (2012)	122	62.30 (8.40)	48-91	65.00	"Predominantly Caucasian"	11.30	9
	Bettcher et al. (2016)	202	73.68 (6.60)	63-99	50.50		17.67	8
	de Frias et al. (2009): CE group[+]	77	66.05 (7.83)	55-86	67.54[7]		15.81	7
	de Frias et al. (2009): CN group[+]	276	68.45 (8.65)	54-88	67.54[7]		15.29	
	Frazier et al. (2015)	119	73.00 (6.50)	55-99	54.00		17.50	1
	Hedden & Yoon (2006)[+]	121	72.24 (4.28)	63-82	57.02		15.69	8

Age Group ($\bar{x}$ age)	Author	*N*	Age (years): $\bar{x}$ (*SD*)	Age Range (years)	% Female	% White or Category	Education: $\bar{x}$ (years) or Category	Study Quality[*]
	Hull et al. (2008)[+]	100	60.24 (5.58)	51-74	80.00		25%-High School; 2% -Associates; 49% - Bachelor's; 24% - Advanced degree	7
	Vaughan & Giovanello (2010)	95	74.40 (6.40)	60-90	56.00	85.00	16.10	8
Multiple Ages	Huizinga et al. (2006)[8]	71	7.2	6-8	54.93		0.56	4
		108	11.2	10-12	57.41		3.92	
		111	15.3	14-16	52.25		7.2	
		94	20.8	18-26	76.6		10.55	
	Pettigrew & Martin (2014)[9]	102	21.00 (3.10)	18-32			14.00	10
		60	71.00 (5.00)	64-87			16.00	

Note. CE = Cognitively Elite; CN = Cognitively Normal; LSES = Low Socioeconomic Status Group; MSES = Medium Socioeconomic Status Group.
[*]Study Quality based on items listed under Data Extraction subheading in the Methods section.
[+]Indicates inclusion in the re-analysis.
[1]Arán-Filippetti (2013) reported confirmatory factor analyses for two separate groups (LSES and MSES).
[2]Lambek & Shevlin (2011) reported two separate confirmatory factor analyses for child and adolescent groups.
[3]Lee et al. (2013) provided a far more comprehensive span of ages; however, due to its sequential cohort-design, there was significant overlap between participants at different ages. In order to ensure that the same individuals were not represented twice in the systematic review, and to increase comparability with other designs, only the cross-section with the greatest amount of participants is considered in the current review and presented in the current table. There is a significant amount of demographic information provided in the original article for the cohorts at baseline; however, the data was not available for the cohort selected for consideration in the current review.
[4]Demographic statistics for Rose et al. (2012) reported separately for full-term and pre-term participants, but only one confirmatory factor analysis was run using the full sample. Some statistics were pulled from Rose et al. (2011), which used the same participant sample.
[5]Xu et al. (2013) reported three separate confirmatory factor analyses for two child and one adolescent group.
[6]Klauer et al. (2010) reported two separate studies, involving separate samples and separate confirmatory factor analyses.
[7]de Frias et al. (2009) did not report separate gender breakdowns for their CE and CN subgroups, so the value reported above was from full sample.
[8]Huizinga et al. (2006) reported demographics for four separate age groups, and reported fit indices corresponding to a configural invariance model across age groups.
[9]Pettigrew & Martin (2014) merged their young and old participants into one group for their confirmatory factor analysis, but did not report separate demographic characteristics for the merged group.

Table 2. *Studies Reporting Measurement Models of Executive Functions: Fit Indices and Latent Constructs*

Age Group	Author	χ^2 (*p*)	*df*	CFI	RMSEA	$\hat{\pi}^1$	Accepted Model	EF-related Factors	Factors tested, but removed/merged	Non-EF Factors in Model
Preschool (<6 yrs.)	Carlson et al. (2014)[2]	24.25 (0.15); 24.56 (0.14)	18; 18	0.96; 0.96	0.06; 0.06	0.15	Both One- and Two-Factor Acceptable	EF; Conflict EF, Delay EF	Conflict and Delay EF merged to form EF	None
	Lerner & Lonigan (2014)	55.60*	33	0.97	0.05	0.78	Two-Factor	Inhibitory Control, WM	Inhibitory Control - Suppression and Inhibitory Control - Conflict merged to form Inhibitory Control	None
	Masten et al. (2012)			0.97	0.04		One-Factor	EF	Hot EF and Cool EF merged to form EF	IQ
	Miller et al. (2012)[+]	43.41	42	1.00	0.02	0.39	Two-Factor	Inhibition, WM	WM and Set-Shifting merged to form WM	None
	Monette et al. (2015)	60.92 (0.59)	64	1.00	0.00	0.88	Two-Factor	Inhibition, Flexibility-WM	Flexibility and WM merged to form Flexibility-WM	Speed (Control Factor)
	Usai et al. (2014)[+]	9.48	8	0.98	0.03	0.20	Two-Factor	Inhibition, WM-Shifting	WM and Shifting merged to form WM-Shifting	None
	Wiebe et al. (2008)	31.14 (0.27)	27	0.99	0.03	0.52	One-Factor	EF	WM, Interference from Distractors, and Proactive Interference merged to form EF	None
	Wiebe et al. (2011)	14.84 (0.39)	14	0.99	0.02	0.36	One-Factor	EF	Inhibition and WM and merged to form EF	None
	Willoughby et al. (2012a)	6.30 (0.71)	9	1.00	0.00	0.96	One-Factor	EF	Inhibitory Control/Attention Shifting and WM merged to form EF	None

Age Group	Author	χ^2 (*p*)	*df*	CFI	RMSEA	$\hat{\pi}$[1]	Accepted Model	EF-related Factors	Factors tested, but removed/merged	Non-EF Factors in Model
School-Aged (6-12 yrs.)	Agostino et al. (2010)[+]	33.26 (0.23)	28	0.98	0.035	0.32	Three-Factor	Inhibition, Updating, Shifting	None	Mental-Attentional Capacity
	Arán-Filippetti (2013)[3]: LSES[+]	30.65 (0.13)	23	0.97	0.05	0.18	Three-Factor	Inhibition, WM, Cognitive Flexibility	None	None
	Arán-Filippetti (2013)[3]: MSES[+]	21.35 (0.56)	23	1.00	0.00	0.18	Three-Factor	Inhibition, WM, Cognitive Flexibility	None	None
	Brocki & Tillman (2014)	16.78 (0.61)	19	1.00	<.001	0.15	Two-Factor	Inhibition, WM	None	None
	Brydges et al. (2012)[+]	20.11 (0.45)	20	1.00	0.01	0.33	One-Factor	EF	Inhibition, Shifting and WM merged to form EF	None
	Duan et al. (2010)[4+]	8.04 (0.24)		0.98	0.08		Three-Factor	Inhibition, Updating, Shifting	None	None
	Lambek & Shevlin (2011)[5]	3.07 (0.80)	6	1.00	0.00	0.10	Three-Factor	Inhibition, Verbal WM, Visuospatial WM	None	None
	Lee et al. (2012)	135.14*	86	0.94	0.06	0.66	Two-factor	Inhibition/Switch, Updating	Inhibition and Switching merged to form Inhibition/Switch	Reaction Time, Flanker Task, Simon Task (All Control Factors)
	Lee et al. (2013)[6]	145.47*	68	0.97	0.06	0.99	Two-Factor	Inhibition/Switch, Updating	Inhibition and Switching merged to form Inhibition/Switch	Control conditions for each task predicted indicators from the same task
	Lehto et al. (2003)[+]	13.73	16	1.00		0.13	Three-Factor	Inhibition, WM, Shifting	None	None

Age Group	Author	χ^2 (*p*)	*df*	CFI	RMSEA	$\hat{\pi}$[1]	Accepted Model	EF-related Factors	Factors tested, but removed/merged	Non-EF Factors in Model
	Rose et al. (2012)[+]	41.88 (0.11)	32	0.96	0.05	0.32	Three-Factor	Inhibition, WM, Shifting	WM: Storage and WM: Storage and Processing merged to form WM	None
	van der Sluis et al. (2007)	190.99*	122	0.95	0.05	0.78	Nested Factor Model	Updating, Shifting	Inhibition merged with Naming	Naming (Control factor)
	van der Ven et al. (2013)	173.43* (0.00)	121	0.96	0.05	0.93	Two-Factor	Inhibition/Shifting, Updating	Inhibition and Shifting merged to form Inhibition/Shifting	Verbal Speed, Motor Speed (Both Control Factors)
	Xu et al. (2013)[7]: Ages 7-9[+]	15.65 (0.34)	14	0.95	0.03	0.19	One-Factor	EF	Inhibition, Updating WM, and Shifting merged to form EF	
	Xu et al. (2013)[7]: Ages 10-12[+]	19.24 (0.30)	14	0.95	0.05	0.19	One Factor	EF	Inhibition, Updating WM, and Shifting merged to form EF	
Adolescents (13-17 yrs.)	Friedman et al. (2011)[+]	53.56*	21	0.96	0.04	0.99	Nested Factor Model	EF, Updating, Shifting	Inhibition merged with EF	None
	Lambek & Shevlin (2011)[5]	3.99 (0.67)	6	1.00	0.00	0.08	Three-Factor	Inhibition, Verbal WM, Visuospatial WM	None	None
	Xu et al. (2013)[7]: Ages 13 to 15[+]	15.72 (0.15)	11	0.95	0.05	0.15	Three-Factor	Inhibition, Updating WM, Shifting	None	None
Adults (18-59 yrs.)	Chuderski et al. (2012)	70.20	60	0.96	0.02	0.52	Four-Factor	Attention Control, Interference Resolution, Response Inhibition, Storage Capacity	Updating merged with Storage Capacity	None
	Fleming et al. (2016)[8+]	30.36 (0.09)	21	0.97	0.03	0.74	Nested Factor Model	EF, Updating, Shifting	Inhibition not tested, but indicators included on EF	None

Age Group	Author	χ^2 (p)	*df*	CFI	RMSEA	$\hat{\pi}$[1]	Accepted Model	EF-related Factors	Factors tested, but removed/merged	Non-EF Factors in Model
									bifactor	
	Fournier-Vicente et al. (2008)	91.62 (0.17)	80	0.99	0.03	0.83	Five-Factor	Verbal SPC, Visuospatial SPC, Selective Attention, Shifting, Strategic Retrieval	Dual-Task Coordination removed	None
	Ito et al. (2015)[+]	32.01 (0.04)*	20	0.98	0.04	0.87	Nested Factor Model	EF, Updating, Shifting	Inhibition not tested, but indicators included on EF bifactor	None
	Klauer et al. (2010)[9] - Study 1[+]	12.82 (0.38)	12	0.98	0.02	0.15	Two-Factor	Inhibition/WM, Switching	Inhibition merged with WM to form Inhibition/WM	None
	Klauer et al. (2010)[9] - Study 2[+]	41.09 (0.13)	32	0.94	0.05	0.20	Three-Factor	Inhibition, WM, Switching	None	None
	Miyake et al. (2000)[+]	20.29 (0.65)	24	1.00		0.28	Three-Factor	Inhibition, Updating, Shifting	None	None
	Was (2007)	12.61 (0.13)	8	0.97	0.06	0.20	Two-Factor	Inhibiting, Updating	None	None
Older Adults (>60 yrs.)	Adrover-Roig et al. (2012)	19.86 (0.70)	24	1.00	0.00	0.18	Three-Factor	WM, Shifting, Access	Inhibition and Updating merged to form WM	None
	Bettcher et al. (2016)	144.12*	95	0.96	0.05	0.88	Two-Factor	Shifting/Inhibition, Updating/WM	Mental Set-shifting and Inhibition merged to form Shifting/Inhibition	Speed (Control Factor)
	de Frias et al. (2009)[10]: CE Group[+]	6.53 (0.69); 5.55 (0.48)	9; 6	1.00; 0.98	0.00; 0.00	0.11	Three-Factor	Inhibition, Updating, Shifting	None	None
	de Frias et al. (2009)[10]: CN Group[+]	17.11 (0.05); 5.11 (0.53)	9; 6	0.94; 1.00	0.06; 0.00	0.32	Two-factor	Not specified for untested two-factor model	None	None

Age Group	Author	χ^2 (p)	*df*	CFI	RMSEA	$\hat{\pi}$[1]	Accepted Model	EF-related Factors	Factors tested, but removed/merged	Non-EF Factors in Model
	Frazier et al. (2015)						Two-Factor	WM, Cognitive Control	Inhibition merged with Set-Shifting to form Cognitive Control	Processing Speed (Control)
	Hedden & Yoon (2006)[+]	115.09	125	1.00	0.00	0.50	Two-Factor	Shifting/Updating, Resistance to Proactive Interference	Shifting and Updating merged to form Shifting/Updating; Prepotent Response Inhibition merged with Speed	Verbal Memory, Visual Memory, Speed
	Hull et al. (2008)[+]	17.76	14		0.05	0.13	Two-Factor	Updating, Shifting	Inhibition removed	None
	Vaughan & Giovanello (2010)	30.23 (0.18)	24	0.97	0.05	0.18	Three-factor	Inhibition, Updating, Task Switching	None	None
Multiple Ages	Huizinga et al. (2006)[11]	139.34*	67				Two-Factor	Stop-Signal Inhibition, Eriksen Flanker Inhibition, Stroop Inhibition, WM, Shifting	Inhibition split into three single-item factors	Basic Speed (Control)
	Pettigrew & Martin (2014)[12]	55.23	42	0.95	0.04		Two-Factor	WM, Interference Resolution	Response-distractor Inhibition and Resistance to Proactive Interference Merged to form Interference Resolution	Age (Control)

Note. CE = Cognitively Elite; CFI = Comparative Fit Index; CN = Cognitively Normal; EF = Executive Function; LSES = Low Socioeconomic Status Group; MSES = Medium Socioeconomic Status Group; RMSEA = Root Mean Square Error of Approximation; SPC = Storage and Processing Coordination; WM = Working Memory; $\hat{\pi}$ = Estimated Power.
*Indicates a significant χ^2 test of model fit (p<.05). [+]Indicates inclusion in the re-analysis.

[1]Power was estimated based on tables provided by Hancock (2006) for post-hoc power analyses of model fit. The values provided herein were based on tables for models with an RMSEA=.02. Because Hancock provided *n* or *df* values in increments of 50 and 5, respectively, the *n* and *df* values from the studies included in the systematic review were rounded to the nearest increments. For models that did not report their *df*, a power value was not estimated, and the studies reporting these models were provided no points for the power criterion of the study quality scale.
[2]Carlson et al. (2014) did not report a preference for either their one- or two-factor model, and the results for both models are reported here, with the one-factor fit indices coming before the semicolon and the two-factor fit indices coming after the semicolon.
[3]Arán-Filippetti (2013) reported confirmatory factor analyses for two separate groups (LSES and MSES).
[4]Duan et al. (2010) reported a χ^2/df value, which is reported here in place of a χ^2 value.
[5]Lambek & Shevlin (2011) reported two separate confirmatory factor analyses for child and adolescent groups.
[6]Lee et al. (2013) used a sequential cohort design, where they recruited participant at different baseline ages and assessed them longitudinally over the course of four years. Consequently, the summary data and fit indices provided for each age group involved a variable amount of overlap (e.g., children starting at age 8 were combined with children that started at age 5 that had already completed three past annual waves of data collection). In turn, only the fit indices from the time point with the largest amount of participants was considered to avoid representing the same individuals twice in the analyses.
[7]Xu et al. (2013) reported three separate confirmatory factor analyses for two child and one adolescent group.
[8]Fleming et al. (2016) and Ito et al. (2015) both included indicators for an inhibition factor, but had these indicators load directly on a general EF bifactor. Guided by previous research, these authors never tested a model including a specific inhibition factor; but because the model included these indicators, inhibition is listed as a factor tested, but removed/merged.
[9]Klauer et al. (2010) reported two separate studies, involving separate samples and separate confirmatory factor analyses.
[10]de Frias et al. (2009) concluded that the data supported a three-factor solution for the CE group (based partially on longitudinal invariance testing); however, the one-factor model fit the data better at Wave 1, and this model was also more parsimonious. The authors concluded a two-factor model best fit the data CN group, although such a model was not tested by the researchers. The fit indices reported herein derive from the one-factor and three-factor models at Wave 1 in their longitudinal design, with the one-factor fit indices coming before the semicolon and the three-factor fit indices coming after the semicolon.
[11]Huizinga et al. (2006) reported fit indices corresponding to a configural invariance model across age groups.
[12]Pettigrew & Martin (2014) merged their young and old participants into one group for their confirmatory factor analysis.

Table 3. *Counts and Frequencies of Constructs represented in Accepted Measurement Models*

	Age Group (Age range)	*k*	EF	Inhibition	UWM	Shifting	Inhibition/Shifting	Inhibition/UWM	Shifting/UWM	SR/Access
Counts (*k*)	All Ages	45	11	23	32	20	5	1	3	2
	Preschool (<6 yrs.)	9	5	4	2	0	0	0	2	0
	School-Aged (6-12 yrs.)	15	3	8	12	7	3	0	0	0
	Adolescents (13-17 yrs.)	3	1*	2	3	2	0	0	0	0
	Adults (18-59 yrs.)	8	2*	5	7	6	0	1	0	1
	Older Adults (>60 yrs.)	8	0	3	6	4	2	0	1	1
Frequencies (%)	All Ages	44	25.00	52.27	72.73	45.45	11.36	2.27	6.82	4.55
	Preschool (<6 yrs.)	9	55.56	44.44	22.22	0.00	0.00	0.00	22.22	0.00
	School-Aged (6-12 yrs.)	15	20.00*	53.33	80.00	46.67	20.00	0.00	0.00	0.00
	Adolescents (13-17 yrs.)	3	33.33*	66.67	100.00	66.67	0.00	0.00	0.00	0.00
	Adults (18-59 yrs.)	8	25.00	62.50	87.50	75.00	0.00	12.50	0.00	12.50
	Older Adults (>60 yrs.)	8	0.00	42.86	85.71	57.14	28.57	0.00	14.29	14.29

Note. * The EF factor observed for adolescent and adult samples were general bifactors in models that also included updating and shifting in the same model. EF = Executive Function; SR = Strategic Retrieval; UWM = Updating/Working Memory. The names attributed to similar constructs differed across studies. Selective Attention, Attention Control, Interference Resolution and Response Inhibition, Resistance to Proactive Interference, Inhibitory Control, Inhibiting and Interference Resolution were subsumed under Inhibition. Updating, WM and Storage Capacity were subsumed under Updating/Working Memory. Cognitive Flexibility, Flexibility, Task Switching and Switching were subsumed by Shifting. Cognitive Control was subsumed under Inhibition/Shifting. Strategic Retrieval and Access were subsumed under Strategic Retrieval/Access. Based on semantic overlap (Packwood et al., 2011), Selective Attention (Fournier-Vicente et al., 2008) and Attention Control (Chuderski et al., 2012) could be subsumed under Shifting; however, the indicators for these factors from both studies were more closely related to Inhibition (e.g., Stroop, Antisaccade), and were thus subsumed under that construct.

Some studies found multiple factors interpretable as sub-dimensions of a common EF-related construct. In these cases, these multiple factors were tallied as representative of a single factor. Specifically, Lambek and Shevlin (2011) found separable Verbal and Visuospatial WM factors, which were tallied as one observation of an Updating/WM factor for each of these authors' two reported samples. Chuderski et al. (2012) found separable Attention Control, Interference Resolution and Response Inhibition factors, which were tallied as one observation of an inhibition factor based on their similarities to this construct based on the authors' conceptual and operational definitions. Lastly, Fournier-Vicente et al. (2008) found separable Verbal and Visuospatial Storage and Processing Coordination, which were tallied as one observation of Updating/WM.

Carlson et al. (2014) did not report a preference for either their one-factor or two-factor model. Based on fit indices, the one-factor was more parsimonious and showed nearly identical fit to the two-factor model. In turn, an EF factor was added to the tally for this study. For the Cognitively Normal group described by de Frias et al. (2009), the authors reported an untested two-factor model as their accepted model. Because this model was untested, it is not clear which factors were represented in this two-factor model, and the results of this group are not represented within this table.

Table 4. *Child and Adolescent Studies: Tests included as Indicators for Executive Function Factors in Accepted Measurement Models*

Author[1]	Inhibition	UWM	Shifting	Inhibition/Shifting
Agostino et al. (2010)	Antisaccade Task	Letter Memory Task	Contingency Naming Task	
	Stroop: Number version	Visual N-Back Task: 0- back, 1-back, 2-back	Trail Making Test	
	Stroop: Color version			
Arán-Filippetti et al. (2013)	Matching Familiar Figures Test	WISC-IV Letter-Number Sequencing	Semantic Verbal Fluency Test	
	Porteus Maze Test	WISC-IV Digit Span Forward	Phonological Verbal Fluency	
	Stroop Color-Word Test	WISC-IV Digit Span Backward	Wisconsin Card Sorting Test: Categories Completed	
Brocki et al. (2014)	Stroop Color-Word Test	Box Task		
	Day-Night Stroop-like Task	WISC-III Digit Span Backward		
	Attention Network Task: Child version	Letter-Number Sequencing		
	Go/No-go Task	Pig House task		
Brydges et al. (2012)	Stroop task: Color-Word	WISC-IV Letter-Number Sequencing	Wisconsin Card Sorting Test: Perseverative Errors	
	Go/No-go Task	WISC-IV Digit Span Backward	Verbal Fluency	
	Compatibility Reaction Time	NEPSY Sentence Repetition	Letter Monitoring	
Duan et al. (2010)	Digit Go/No-go Task	Digit 2-back Task	Odd-More Task: Digit Shifting Task	
	Figure Go/No-go Task	Figure Position 2-back	Local Global: Figure Shifting Task	
Huizinga et al. (2006)[2]	Stop-Signal Task	Tic Tac Toe Task	Local Global Task	
	Eriksen Flanker Task	Mental Counters Task	Dots Triangles Task	
	Stroop Task	Running Memory	Smiling Faces Task	
Lambek & Shevlin (2011)[3]	Stop Task	Digit Span Backward		
	Walk Don’t Walk	Letters Backward		

Author[1]	Inhibition	UWM	Shifting	Inhibition/Shifting
		WISC-III-PI Spatial Span Backward		
		Finger Windows Backward		
Lee et al. (2012)	Flanker Task (RT): Incongruent Condition	Mister X Task	Flanker Task (RT): Switch Condition	
	Simon task (RT): Incongruent Condition	Listening Recall Task	Simon task (RT): Switch Condition	
	Flanker task (Accuracy): Incongruent Condition	Pictorial Updating Task	Picture-Symbol Task (RT): Switch Condition	
	Simon task (Accuracy): Incongruent Condition	Flanker task: Accuracy, Switch Condition	Simon Task (Accuracy): Switch Condition	
			Picture-Symbol task (Accuracy): Switch Condition	
Lee et al. (2013)	Flanker Task: Incongruent Condition	Listening Recall Task	Flanker Task: Switch Condition	
	Simon Task: Incongruent Condition	Mister X Task	Simon Task: Switch Condition	
	Mickey Task: Incongruent Condition	Pictorial Updating Task	Picture-Symbol Task: Switch Condition	
Lehto et al. (2003)	CANTAB Tower of London	NEPSY Auditory Attention and Response Set B	NEPSY Word Fluency Task	
	Matching Familiar Figures Test	CANTAB Spatial Span Task	Trail Making Test B	
		CANTAB Spatial Working Memory Task		
		WISC-R Mazes Task		
Lerner et al. (2014)[4]	Bird and Dragon Task	Word Span Reversed Task		
	Luria's Hand Game Task	Listening Span Task		
	Picture Imitation	Size Ordering Task		
	Block Sorting	Object Span Task		
	Day-Night Stroop			
	Knock-Tap Task			
Miller et al. (2012)	Preschool Continuous Performance Test: Commission Errors and Ratio	Backward Digit Span	Dimensional Change Card Sort: Border Version;	
	Boy-Girl Stroop	Backward Word Span Test	Go/No-Go 2: Hit Ratio	

Author[1]	Inhibition	UWM	Shifting	Inhibition/Shifting
	Tower of Hanoi	Boxes Task: Hit Ratio	Go/No-Go 3: Hit Ratio	
	Go/No-Go: Commission Errors and Ratio	Preschool Continuous Performance Test: Omissions		
Monette et al. (2015)	Fruit Stroop Task Interference: Correct Responses	Backward Word Span: Correct Responses	Verbal Fluidity Shift: Shift Errors	
	Hand Stroop Interference: Correct Responses	Backward Block Span: Correct Responses	Card Sort: Perseverative Errors	
	Day-night Test: Correct Responses	Trails-P Test: Perseverative Errors		
	Fruit Stroop Interference: Errors	Face Sort: Perseverative Errors		
	Hand Stroop Interference: Errors			
	Day-Night Test: Errors			
Rose et al. (2012)[5]	Go/No-go	CANTAB Spatial Span	Trail-Making: Time B - A	
	CANTAB Rapid Visual Information Processing	Pattern Span	CANTAB Intradimensional-Extradimensional Shift: Reversal Trials	
		Change Detection		
		CANTAB Spatial Working Memory		
		Counting Span		
		Listening Span		
Usai et al. (2014)	Circle Drawing Task	Dual Request Selective Task	Semantic Fluency	
	Tower of London Task	Backward Digit Span	Dimensional Change Card Sort	
van der Sluis et al. (2007)	Numerical Size Inhibition	Letter Memory Task	Making Trails Task	
	Object Inhibition Task	Digit Memory Task	Objects Shifting Task	
	Stroop Color-Word Test	Keep Track Task	Place Shifting Task	
	Quantity Inhibition Task		Symbol Shifting Task	
van der Ven (2013)	Animal Stroop Task: Inhibition Accuracy	Digit Span Backwards Task	Animal Shifting Task: Shifting Accuracy	
	Local Global Task: Inhibition Accuracy	Odd One Out Task	Sorting Task: Shifting Accuracy	

Author[1]	Inhibition	UWM	Shifting	Inhibition/Shifting
	Simon Task: Inhibition Accuracy	Keep Track Task	Animal Shifting: Shifting Speed	
	Animal Stroop: Inhibition Speed		Sorting Task: Shifting Speed	
	Local Global: Inhibition Speed		Trail Making Test in Colors: Shifting Accuracy	
	Simon Task: Inhibition Speed		Trail Making Test in Colors: Shifting Speed	
Wiebe et al. (2008)[6]	Child Continuous Performance Test	Six Boxes Test		
	NEPSY Statue Subtest	Delayed Alternation Task		
	Tower of Hanoi Task	Digit Span Subtest: Differential Abilities Scale		
	NEPSY Visual Attention Subtest			
	Delayed Response Task			
	Shape School Inhibit Condition			
	Whisper Task			
Wiebe et al. (2011)	Big-Little Stroop Task	Nebraska Barnyard Task		
	Go/No-go Task	Nine Boxes Task		
	Shape School Task: Inhibit Condition	Delayed Alternation Task		
	Snack Delay Task			
Willoughby et al. (2012)		Working Memory Span		Simon Task: Spatial Conflict Arrows:
		Pick the Picture		Animal Go No-Go Task
				Silly Sounds Stroop Task
				Something's the Same Task
Xu et al. (2013)	Go/No-go	N-back Task: 1-back, 2-back	Number-Pinyin Task	
	Color-Word Stroop Task: RT Ratio	Running Memory Task	Dots-Triangles Task	

Note. CANTAB = Cambridge Neuropsychological Test Automated Battery; EF = Executive Function; RT = Reaction Time; UWM = Updating/Working Memory. WISC = Wechsler Intelligence Scale for Children. The constructs of Updating and Working Memory were subsumed under the column UWM. The constructs of Set-Shifting, Flexibility, Cognitive Flexibility, and Switching were subsumed under the column of Shifting. The construct of Inhibitory Control/Attention Shifting, represented by just Willoughby et al. (2012a) was subsumed under Inhibition/Shifting.

[1]Two studies assigned tests to constructs not represented in the table above. Carlson et al. (2014) assigned Bear/Dragon, Backward Digit Span, Less is More, Grass/Snow, and the Dimensional Change Card Sort to Conflict EF; and Gift Delay, Delay of Gratification, and Tower Building to Delay EF. Masten et al. (2012) assigned Dinky Toys, Gift Delay Part I, and Gift Delay Part II to Hot EF; and Simon Says, Dimensional Change Card Sort, Computerized Pointing Stroop, and Peg Tapping to Cool EF.
[2]Huizinga et al. (2006) had multiple age groups, ranging from 6 to 26, but because the age groups were predominantly child/adolescent, they were included in this table.
[3]Lambek and Shevlin (2011) assigned indicators to both Verbal and Visuospatial Working Memory, and the indicators for both constructs listed under the Inhibition category.
[4]Lerner et al. (2014) assigned indicators to Inhibitory Control – Suppression and Inhibitory Control – Conflict, and the indicators for both constructs were listed under the Inhibition category.
[5]Rose et al. (2012) assigned indicators to Working Memory: Storage and Working Memory: Storage and Processing, and the indicators for both constructs were listed under the UWM category.
[6]Wiebe et al. (2008) assigned indicators to Interference from Distractors and Proactive Interference, and the indicators for both constructs were listed under the Inhibition category.

Table 5. *Adult Studies: Tests included as Indicators for Executive Function Factors in Accepted Measurement Models*

Author	Inhibition	UWM	Shifting	Access/SR
Adrover-Roig et al. (2012)		WAIS-III Digit Forward Subtest	Brixton Test: Errors	Boston Naming Test
		CANTAB Paired Associated Learning: Errors	Madrid Card Sorting Test: Switch Cost	Controlled Oral Word Association Test: FAS
		Stroop Test: Color-Word	Madrid Card Sorting Test: Efficient Series	Semantic Fluency: Animals
Bettcher et al. (2016)	Enclosed Flanker Test: Incongruent Condition	Dot Counting: Total Correct	Set-Shifting Task: Shift Trials	
	Antisaccade Task: Proportion Correct	Running-Letter Memory Task: Proportion Correct	Number-letter Task: Shift Trials	
	Stroop Interference Test: Total Correct	N-Back: 1-Back and 2-Back: D-prime	Design Fluency Task: Shift Condition Accuracy	
		Digit Span Backward: Total Span		
Chuderski et al. (2012)[1]	Figure-Word Task: Latency	Keep-Track Task		
	Color-word Task: Latency	N-Back Task: 2-back: Figural		
	Stop-Signal Task	Two-Array Comparison Task		
	Go/No-go Task	Monitoring Task: Verbal Version		
	Figure-Word Task: Accuracy	Monitoring Task: Figural Version		
	Color-word Task: Accuracy			
	Number Stroop Task: Accuracy			
	Antisaccade Task			
de Frias et al. (2009)	Hayling Sentence Completion Test	Reading Span Task	Brixton Spatial Anticipation Test	
	Stroop Test	Computational Span Task	Color Trails Test	
Fleming et al. (2016)	Antisaccade Task	Keep Track Task	Color-Shape Task	
	Stop Signal Task	Letter Memory Task	Category Switch Task	
	Stroop Color-Naming Task	Spatial 2-Back Task	Number-Letter Task	

Author	Inhibition	UWM	Shifting	Access/SR
Freidman et al. (2011)	Antisaccade Task	Keep Track Task	Number-Letter Task	
	Stop-Signal Task	Letter Memory Task	Color-Shape Task	
	Stroop Task	Spatial 2-Back Task	Category-Switch Task	
Fournier-Vicente et al. (2008)[2]	Stroop-Color Test	Backward Digit Span Task	Plus-minus Task	Random Letter Generation Task
	Stroop-Numerical Test	Verbal Transposed Span Task	Number-Letter Task	Hayling Test
	d2 Target Detection Task	Verbal Arithmetic Span Task	Local-Global Task	Semantic Verbal Fluency Task
		Backward Location Span Task		
		Visuospatial Transposed Span Task		
		Visuospatial Arithmetic Span Task		
Frazier et al. (2015)	Stroop Task	Digit Span Backward Task	Design Fluency Switching Test	
	Enclosed Flanker Task	N-Back: 1-back and 2-back	Number-letter Task	
		Dot Counting Task		
		Running-Letter Memory Task		
Hedden & Yoon (2006)[3]	Antisaccade	Letter Memory Task	Plus-Minus Task	
	Stroop Color Naming	Backward Digit Span Task	Wisconsin Card Sorting Task	
	Excluded Letter Fluency 1	Self-Ordered Pointing	Trail Making Test	
	Excluded Letter Fluency 2			
	Semantic Fluency			
Hull et al. (2008)	Stroop Verbal Task	Keep Track Verbal Task	Local-Global Verbal Task	
	Stroop Nonverbal Task	Keep Track Non-Verbal Task	Local-Global Non-Verbal Task	
	Antisaccade Task	N-Back Verbal Task	Plus-Minus Task	
		N-Back Non-verbal Task		

Author	Inhibition	UWM	Shifting	Access/SR
Ito et al. (2015)	Antisaccade Task	Keep Track	Number-Letter	
	Stop-Signal Task	Letter Memory	Color-Shape	
	Stroop Task	Spatial N-back	Category Switch	
Klauer et al. (2010) – Study 1	Stop-Signal Task	Reading Span Task	Color-Size Task	
	Antisaccade Task	Operation Span Task	Number-Letter Task	
			Plus-Minus Task	
Klauer et al. (2010) – Study 2	Flanker Task	Reading Span Task	Color-Size Task	
	Stroop Task	Operation Span Task	Semantic Switching Task	
	Simon Task	Counting Span Task	Number-Letter Task	
	Antisaccade Task			
Miyake et al. (2000)	Antisaccade Task	Keep Track Task	Plus-Minus Task	
	Stop-signal Task	Tone Monitoring Task	Number-letter Task	
	Stroop Task	Letter Memory Task	Local-global Task	
Pettigrew et al. (2014)[4]	Recent Negatives Task	Automated Operation Span		
	Cued Recall Task	WAIS-R Backwards Digit Span		
	Release from Proactive Interference Task	Sternberg Recognition Task		
	Flanker Task			
	Picture-Word Interference task			
	Non-verbal Stroop Task			
	Stroop Task			
Vaughan & Giovanello (2010)	Stroop Task: RT Correct Incongruent	N-back Task: 1-back, 2-back, and 3-back: Percent Correct	More-Less and Odd-even Task: RT Correct Switch	
	Anticue Task: RT Correct Invalidly Cued	Letter-memory Task: Percent Correct	Number-Letter Task: RT Correct Switch	
	Stop-Signal Task: % Correct	Refreshing Paradigm: Percent Correct	Local-Global Task: RT Correct Switch	

Author	Inhibition	UWM	Shifting	Access/SR
Was (2007)	Number Disengagement Task	Alphabet Working Memory Task		
	Original Stroop Color Task	ABCD Working Memory		
	Number Stroop Task	Numerical Strings Audio Working Memory		

Note. CANTAB = Cambridge Neuropsychological Test Automated Battery; RT = Reaction Time; SR = Strategic Retrieval; UWM = Updating/Working Memory; WAIS = Wechsler Adult Intelligence Scale. Updating, Working Memory, Storage Capacity, Verbal Storage and Processing Coordination, and Visual Storage and Processing Coordination were subsumed under UWM.
Attention Control, Inhibiting, Interference Resolution, Response Inhibition, Response-distractor Inhibition, Resistance to Proactive Interference, and Selective Attention were subsumed under Inhibition. Task Switching, Switching, Mental Set-Shifting, and Set-Shifting were subsumed under Shifting.
[1]Chuderski et al. (2012) included five factors. Presented in the Inhibition column are the indicators for the Interference Resolution, Response Inhibition, and Attention Control factors. Presented in the UWM column are the indicators for the Updating and Storage Capacity factors.
[2]Fournier-Vicente et al. (2008) included six factors. Presented in the Inhibition column are the indicators for Selective Attention. Presented in the UWM column are the indicators for Verbal Storage and Processing Coordination and Visual Storage and Processing Coordination. Their sixth factor was Dual-Task Coordination, which is not presented in this table. The indicators for this factor included Digit Span + Box-Crossing, Location Span + Categorization, Dual-Storage, and Dual-Processing.
[3]Hedden and Yoon (2006) had two inhibition-related factors (i.e., Prepotent Response Inhibition, Resistance to Proactive Interference), with the indicators for these factors both presented in the Inhibition column.
[4]Pettigrew et al. (2014) included two inhibition-related factors (i.e., Response-distractor Inhibition, Resistance to Proactive Interference), with the indicators for these factors both presented in the Inhibition column.

Table 6. *Child and Adolescent Studies: Percent Convergence, Percent Meeting Fit Criteria, and Rate of Model Acceptance for 5,000 Bootstrapped Samples by Measurement Model and Study*

			Percent of Converged Models Meeting Fit Criteria				Rate of Model Acceptance based on Fit Thresholds			
			CFI		RMSEA		CFI		RMSEA	
Model	Study	% Converged	≥.90	≥.95	≤.08	≤.05	≥.90	≥.95	≤.08	≤.05
Bifactor	Agostino et al. (2010)	78%	94%	62%	52%	21%	73%	48%	41%	16%
	Arán-Filippetti (2013)	34%	100%	100%	90%	34%	34%	34%	31%	12%
	Brydges et al. (2012)	3%	88%	50%	80%	34%	3%	2%	2%	1%
	Duan et al. (2010)	52%	23%	4%	1%	1%	12%	2%	1%	1%
	Lehto et al. (2003)	67%	60%	26%	39%	14%	40%	17%	26%	9%
	Miller et al. (2012)	9%	49%	5%	18%	0%	4%	0%	2%	0%
	Rose et al. (2012)	49%	28%	5%	31%	3%	14%	2%	15%	1%
	Usai et al. (2014)	82%	45%	20%	35%	14%	37%	16%	29%	11%
	Xu et al. (2013) - Ages 7-9	44%	83%	59%	76%	48%	37%	26%	33%	21%
	Xu et al. (2013) - Ages 10-12	32%	71%	42%	68%	33%	23%	13%	22%	11%
	Xu et al. (2013) - Ages 13-15	74%	25%	6%	13%	3%	19%	4%	10%	2%
	Median*	49%	71%	42%	52%	21%	23%	13%	22%	9%
	Mean*	48%	64%	39%	50%	21%	28%	16%	20%	8%
Nested Factor	Agostino et al. (2010)	100%	55%	14%	17%	3%	55%	14%	17%	3%
	Arán-Filippetti (2013)	73%	100%	99%	66%	10%	73%	72%	48%	7%
	Brydges et al. (2012)	8%	81%	42%	87%	35%	6%	3%	7%	3%
	Duan et al. (2010)	56%	12%	2%	1%	0%	7%	1%	1%	0%
	Lehto et al. (2003)	77%	45%	17%	29%	9%	35%	13%	22%	7%
	Miller et al. (2012)	14%	3%	0%	1%	0%	0%	0%	0%	0%
	Rose et al. (2012)	60%	25%	4%	30%	3%	15%	2%	18%	2%
	Usai et al. (2014)	94%	21%	8%	21%	6%	20%	8%	20%	6%
	Xu et al. (2013) - Ages 7-9	46%	83%	59%	80%	51%	38%	27%	37%	23%
	Xu et al. (2013) - Ages 10-12	34%	75%	46%	76%	39%	26%	16%	26%	13%
	Xu et al. (2013) - Ages 13-15	76%	29%	8%	19%	5%	22%	6%	14%	4%
	Median*	60%	55%	17%	30%	9%	26%	13%	18%	4%
	Mean*	59%	56%	32%	45%	17%	31%	17%	21%	7%

Model	Study	% Converged	Percent of Converged Models Meeting Fit Criteria: CFI ≥.90	CFI ≥.95	RMSEA ≤.08	RMSEA ≤.05	Rate of Model Acceptance based on Fit Thresholds: CFI ≥.90	CFI ≥.95	RMSEA ≤.08	RMSEA ≤.05
Three-Factor	Agostino et al. (2010)	34%	94%	65%	65%	30%	32%	22%	22%	10%
	Arán-Filippetti (2013)	65%	92%	7%	0%	0%	60%	5%	0%	0%
	Brydges et al. (2012)	2%	77%	25%	93%	23%	2%	1%	2%	0%
	Duan et al. (2010)	21%	28%	7%	1%	0%	6%	1%	0%	0%
	Lehto et al. (2003)	86%	64%	32%	52%	22%	55%	28%	45%	19%
	Miller et al. (2012)	86%	4%	0%	2%	0%	3%	0%	2%	0%
	Rose et al. (2012)	60%	2%	0%	5%	0%	1%	0%	3%	0%
	Usai et al. (2014)	24%	56%	30%	45%	19%	13%	7%	11%	5%
	Xu et al. (2013) - Ages 7-9	15%	64%	38%	66%	33%	10%	6%	10%	5%
	Xu et al. (2013) - Ages 10-12	14%	72%	42%	77%	40%	10%	6%	11%	6%
	Xu et al. (2013) - Ages 13-15	26%	37%	13%	32%	10%	10%	3%	8%	3%
	Median*	26%	64%	25%	52%	22%	10%	5%	8%	3%
	Mean*	36%	59%	25%	43%	18%	21%	8%	11%	5%
Inhibition-Shifting Merged	Agostino et al. (2010)	47%	58%	16%	23%	4%	27%	8%	11%	2%
	Arán-Filippetti (2013)	100%	83%	3%	0%	0%	83%	3%	0%	0%
	Brydges et al. (2012)	89%	55%	14%	79%	15%	49%	12%	70%	13%
	Duan et al. (2010)	33%	11%	2%	1%	0%	4%	1%	0%	0%
	Lehto et al. (2003)	98%	41%	16%	38%	12%	40%	16%	37%	12%
	Miller et al. (2012)	88%	0%	0%	0%	0%	0%	0%	0%	0%
	Rose et al. (2012)	92%	0%	0%	2%	0%	0%	0%	2%	0%
	Usai et al. (2014)	35%	24%	10%	28%	9%	8%	4%	10%	3%
	Xu et al. (2013) - Ages 7-9	58%	56%	33%	67%	32%	32%	19%	39%	19%
	Xu et al. (2013) - Ages 10-12	69%	64%	35%	76%	38%	44%	24%	52%	26%
	Xu et al. (2013) - Ages 13-15	98%	28%	8%	31%	6%	27%	8%	30%	6%
	Median*	89%	55%	14%	31%	6%	32%	8%	30%	6%
	Mean*	76%	44%	14%	35%	12%	34%	10%	27%	9%
Inhibition-Updating Merged	Agostino et al. (2010)	59%	58%	16%	23%	4%	34%	9%	14%	2%
	Arán-Filippetti (2013)	64%	74%	1%	0%	0%	47%	1%	0%	0%
	Brydges et al. (2012)	6%	61%	16%	80%	16%	4%	1%	5%	1%
	Duan et al. (2010)	35%	4%	1%	0%	0%	1%	0%	0%	0%
	Lehto et al. (2003)	94%	36%	12%	33%	9%	34%	11%	31%	8%

Model	Study	% Converged	Percent of Converged Models Meeting Fit Criteria: CFI ≥.90	CFI ≥.95	RMSEA ≤.08	RMSEA ≤.05	Rate of Model Acceptance based on Fit Thresholds: CFI ≥.90	CFI ≥.95	RMSEA ≤.08	RMSEA ≤.05
	Miller et al. (2012)	93%	0%	0%	0%	0%	0%	0%	0%	0%
	Rose et al. (2012)	87%	1%	0%	4%	0%	1%	0%	3%	0%
	Usai et al. (2014)	52%	26%	11%	29%	10%	14%	6%	15%	5%
	Xu et al. (2013) - Ages 7-9	35%	57%	32%	66%	32%	20%	11%	23%	11%
	Xu et al. (2013) - Ages 10-12	40%	63%	37%	76%	39%	25%	15%	30%	16%
	Xu et al. (2013) - Ages 13-15	78%	13%	3%	16%	2%	10%	2%	12%	2%
	Median*	59%	57%	12%	23%	4%	20%	2%	12%	2%
	Mean*	55%	41%	13%	33%	11%	20%	6%	13%	4%
Shifting-Updating Merged	Agostino et al. (2010)	99%	93%	61%	69%	30%	92%	60%	68%	30%
	Arán-Filippetti (2013)	100%	88%	4%	0%	0%	88%	4%	0%	0%
	Brydges et al. (2012)	54%	50%	12%	75%	12%	27%	6%	41%	6%
	Duan et al. (2010)	44%	5%	0%	0%	0%	2%	0%	0%	0%
	Lehto et al. (2003)	97%	37%	13%	33%	9%	36%	13%	32%	9%
	Miller et al. (2012)	98%	1%	0%	1%	0%	1%	0%	1%	0%
	Rose et al. (2012)	90%	0%	0%	2%	0%	0%	0%	2%	0%
	Usai et al. (2014)	75%	50%	25%	52%	22%	38%	19%	39%	17%
	Xu et al. (2013) - Ages 7-9	56%	57%	31%	65%	31%	32%	17%	36%	17%
	Xu et al. (2013) - Ages 10-12	51%	61%	33%	75%	35%	31%	17%	38%	18%
	Xu et al. (2013) - Ages 13-15	51%	12%	3%	14%	2%	6%	2%	7%	1%
	Median*	56%	50%	12%	33%	9%	31%	6%	32%	6%
	Mean*	71%	45%	17%	37%	13%	35%	13%	25%	9%
Unidimensional	Agostino et al. (2010)	100%	57%	16%	27%	5%	57%	16%	27%	5%
	Arán-Filippetti (2013)	100%	71%	1%	0%	0%	71%	1%	0%	0%
	Brydges et al. (2012)	100%	48%	11%	77%	14%	48%	11%	77%	14%
	Duan et al. (2010)	74%	2%	0%	0%	0%	1%	0%	0%	0%
	Lehto et al. (2003)	100%	21%	6%	21%	5%	21%	6%	21%	5%
	Miller et al. (2012)	99%	0%	0%	0%	0%	0%	0%	0%	0%
	Rose et al. (2012)	100%	0%	0%	1%	0%	0%	0%	1%	0%
	Usai et al. (2014)	95%	24%	10%	34%	10%	23%	10%	32%	10%
	Xu et al. (2013) - Ages 7-9	100%	55%	31%	67%	32%	55%	31%	67%	32%
	Xu et al. (2013) - Ages 10-12	100%	60%	33%	77%	37%	60%	33%	77%	37%

			Percent of Converged Models Meeting Fit Criteria				Rate of Model Acceptance based on Fit Thresholds			
			CFI		RMSEA		CFI		RMSEA	
Model	Study	% Converged	≥.90	≥.95	≤.08	≤.05	≥.90	≥.95	≤.08	≤.05
	Xu et al. (2013) - Ages 13-15	100%	12%	3%	17%	2%	12%	3%	17%	2%
	Median*	100%	48%	6%	21%	5%	48%	6%	21%	5%
	Mean*	97%	36%	11%	32%	11%	36%	11%	32%	11%

Note. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation. *Median and Mean values exclude the two preschool articles listed in the table (Miller et al., 2012; Usai et al., 2014).

Table 7. *Adult Studies: Percent Convergence, Percent Meeting Fit Criteria, and Rate of Model Acceptance for 5,000 Bootstrapped Samples by Measurement Model and Study*

			Percent of Bootstrapped Samples Meeting Fit Cutoffs				Rate of Model Acceptance based on Fit Thresholds			
			CFI		RMSEA		CFI		RMSEA	
Model	Study	% Converged	≥.90	≥.95	≤.08	≤.05	≥.90	≥.95	≤.08	≤.05
Bifactor	de Frias et al. (2009) - CE	68%	52%	35%	42%	28%	35%	24%	29%	19%
	de Frias et al. (2009) - CN	98%	90%	59%	79%	39%	88%	58%	77%	38%
	Fleming et al. (2016)	8%	95%	46%	98%	39%	8%	4%	8%	3%
	Friedman et al. (2011)	13%	100%	88%	100%	43%	13%	11%	13%	6%
	Hedden & Yoon (2006)	5%	38%	15%	35%	13%	2%	1%	2%	1%
	Hull et al. (2008)	6%	0%	0%	0%	0%	0%	0%	0%	0%
	Ito et al. (2015)	6%	99%	59%	96%	26%	6%	4%	6%	2%
	Klauer et al. (2010) - Study 1	21%	42%	21%	51%	23%	9%	4%	11%	5%
	Klauer et al. (2010) - Study 2	5%	54%	14%	45%	9%	3%	1%	2%	0%
	Miyake et al. (2000)	13%	55%	24%	68%	24%	7%	3%	9%	3%
	Median	11%	55%	30%	60%	25%	7%	4%	8%	3%
	Mean	24%	63%	36%	61%	24%	17%	11%	16%	8%
Nested Factor	de Frias et al. (2009) - CE	71%	56%	39%	49%	32%	40%	28%	35%	23%
	de Frias et al. (2009) - CN	98%	92%	64%	88%	51%	90%	63%	86%	50%
	Fleming et al. (2016)	89%	81%	22%	97%	27%	72%	20%	86%	24%
	Friedman et al. (2011)	22%	74%	1%	63%	0%	16%	0%	14%	0%
	Hedden & Yoon (2006)	5%	35%	12%	41%	11%	2%	1%	2%	1%
	Hull et al. (2008)	25%	0%	0%	0%	0%	0%	0%	0%	0%
	Ito et al. (2015)	96%	87%	14%	86%	6%	84%	13%	83%	6%
	Klauer et al. (2010) - Study 1	22%	47%	27%	59%	28%	10%	6%	13%	6%
	Klauer et al. (2010) - Study 2	34%	10%	1%	15%	0%	3%	0%	5%	0%
	Miyake et al. (2000)	34%	33%	13%	57%	16%	11%	4%	19%	5%
	Median	34%	52%	14%	58%	14%	14%	5%	17%	6%
	Mean	50%	52%	19%	56%	17%	33%	13%	34%	11%
Three-Factor	de Frias et al. (2009) - CE	7%	69%	48%	56%	39%	5%	3%	4%	3%
	de Frias et al. (2009) - CN	37%	93%	68%	84%	49%	34%	25%	31%	18%
	Fleming et al. (2016)	94%	56%	7%	95%	16%	53%	7%	89%	15%

Model	Study	% Converged	Percent of Bootstrapped Samples Meeting Fit Cutoffs				Rate of Model Acceptance based on Fit Thresholds			
			CFI		RMSEA		CFI		RMSEA	
			≥.90	≥.95	≤.08	≤.05	≥.90	≥.95	≤.08	≤.05
	Friedman et al. (2011)	82%	100%	37%	100%	20%	82%	30%	82%	16%
	Hedden & Yoon (2006)	29%	15%	3%	25%	4%	4%	1%	7%	1%
	Hull et al. (2008)	17%	0%	0%	0%	0%	0%	0%	0%	0%
	Ito et al. (2015)	52%	83%	11%	93%	7%	43%	6%	48%	4%
	Klauer et al. (2010) - Study 1	12%	31%	11%	44%	16%	4%	1%	5%	2%
	Klauer et al. (2010) - Study 2	43%	7%	0%	14%	0%	3%	0%	6%	0%
	Miyake et al. (2000)	83%	24%	9%	57%	14%	20%	7%	47%	12%
	Median	40%	44%	10%	57%	15%	12%	5%	19%	3%
	Mean	46%	48%	19%	57%	17%	25%	8%	32%	7%
Inhibition-Shifting Merged	de Frias et al. (2009) - CE	43%	62%	44%	59%	41%	27%	19%	25%	18%
	de Frias et al. (2009) - CN	91%	83%	49%	84%	41%	76%	45%	76%	37%
	Fleming et al. (2016)	98%	0%	0%	23%	0%	0%	0%	23%	0%
	Friedman et al. (2011)	100%	24%	0%	55%	0%	24%	0%	55%	0%
	Hedden & Yoon (2006)	67%	12%	3%	26%	5%	8%	2%	17%	3%
	Hull et al. (2008)	60%	0%	0%	0%	0%	0%	0%	0%	0%
	Ito et al. (2015)	100%	0%	0%	0%	0%	0%	0%	0%	0%
	Klauer et al. (2010) - Study 1	27%	3%	1%	14%	2%	1%	0%	4%	1%
	Klauer et al. (2010) - Study 2	80%	1%	0%	4%	0%	1%	0%	3%	0%
	Miyake et al. (2000)	94%	4%	1%	27%	3%	4%	1%	25%	3%
	Median	86%	4%	1%	25%	1%	2%	0%	20%	0%
	Mean	76%	19%	10%	29%	9%	14%	7%	23%	6%
Inhibition-Updating Merged	de Frias et al. (2009) - CE	33%	45%	28%	42%	25%	15%	9%	14%	8%
	de Frias et al. (2009) - CN	94%	83%	50%	84%	42%	78%	47%	79%	39%
	Fleming et al. (2016)	100%	4%	0%	58%	0%	4%	0%	58%	0%
	Friedman et al. (2011)	77%	0%	0%	0%	0%	0%	0%	0%	0%
	Hedden & Yoon (2006)	60%	12%	3%	26%	4%	7%	2%	16%	2%
	Hull et al. (2008)	24%	0%	0%	0%	0%	0%	0%	0%	0%
	Ito et al. (2015)	100%	39%	1%	72%	1%	39%	1%	72%	1%
	Klauer et al. (2010) - Study 1	44%	12%	4%	30%	8%	5%	2%	13%	4%
	Klauer et al. (2010) - Study 2	56%	0%	0%	1%	0%	0%	0%	1%	0%
	Miyake et al. (2000)	98%	13%	4%	47%	8%	13%	4%	46%	8%

			Percent of Bootstrapped Samples Meeting Fit Cutoffs				Rate of Model Acceptance based on Fit Thresholds			
			CFI		RMSEA		CFI		RMSEA	
Model	Study	% Converged	≥.90	≥.95	≤.08	≤.05	≥.90	≥.95	≤.08	≤.05
	Median	69%	12%	2%	36%	3%	6%	1%	15%	2%
	Mean	69%	21%	9%	36%	9%	16%	6%	30%	6%
Shifting-Updating Merged	de Frias et al. (2009) - CE	35%	43%	28%	40%	25%	15%	10%	14%	9%
	de Frias et al. (2009) - CN	50%	40%	12%	42%	9%	20%	6%	21%	5%
	Fleming et al. (2016)	97%	0%	0%	0%	0%	0%	0%	0%	0%
	Friedman et al. (2011)	100%	27%	0%	58%	0%	27%	0%	58%	0%
	Hedden & Yoon (2006)	62%	13%	4%	28%	5%	8%	2%	17%	3%
	Hull et al. (2008)	54%	0%	0%	0%	0%	0%	0%	0%	0%
	Ito et al. (2015)	69%	0%	0%	0%	0%	0%	0%	0%	0%
	Klauer et al. (2010) - Study 1	40%	1%	0%	8%	0%	0%	0%	3%	0%
	Klauer et al. (2010) - Study 2	75%	0%	0%	1%	0%	0%	0%	1%	0%
	Miyake et al. (2000)	93%	4%	1%	27%	3%	4%	1%	25%	3%
	Median	66%	3%	0%	18%	0%	2%	0%	9%	0%
	Mean	68%	13%	5%	20%	4%	7%	2%	14%	2%
Unidimensional	de Frias et al. (2009) - CE	76%	40%	26%	40%	25%	30%	20%	30%	19%
	de Frias et al. (2009) - CN	98%	39%	12%	51%	11%	38%	12%	50%	11%
	Fleming et al. (2016)	100%	0%	0%	0%	0%	0%	0%	0%	0%
	Friedman et al. (2011)	100%	0%	0%	0%	0%	0%	0%	0%	0%
	Hedden & Yoon (2006)	100%	10%	3%	26%	5%	10%	3%	26%	5%
	Hull et al. (2008)	93%	0%	0%	0%	0%	0%	0%	0%	0%
	Ito et al. (2015)	100%	0%	0%	0%	0%	0%	0%	0%	0%
	Klauer et al. (2010) - Study 1	81%	0%	0%	6%	0%	0%	0%	5%	0%
	Klauer et al. (2010) - Study 2	98%	0%	0%	0%	0%	0%	0%	0%	0%
	Miyake et al. (2000)	100%	2%	0%	17%	1%	2%	0%	17%	1%
	Median	99%	0%	0%	3%	0%	0%	0%	2%	0%
	Mean	95%	9%	4%	14%	4%	8%	3%	13%	4%

Note. CE = Cognitively Elite; CFI = Comparative Fit Index; CI = Confidence Interval; CN = Cognitively Normal; RMSEA = Root Mean Square Error of Approximation.

Table 8. *Child and Adolescents Studies: Mean Fit Indices (95% CIs) for Converged Models by Measurement Model and Study*

Model	Study	% Converged	CFI	95% CI	RMSEA	95% CI
Bifactor	Agostino et al. (2010)	78%	0.96	(0.88, 1)	0.08	(0, 0.14)
	Arán-Filippetti (2013)	34%	0.99	(0.97, 1)	0.06	(0, 0.09)
	Brydges et al. (2012)	3%	0.94	(0.88, 1)	0.06	(0.02, 0.09)
	Duan et al. (2010)	52%	0.85	(0.73, 0.96)	0.23	(0.12, 0.33)
	Lehto et al. (2003)	67%	0.91	(0.77, 1)	0.09	(0, 0.15)
	Miller et al. (2012)	9%	0.90	(0.83, 0.96)	0.10	(0.06, 0.13)
	Rose et al. (2012)	49%	0.87	(0.74, 0.96)	0.09	(0.05, 0.13)
	Usai et al. (2014)	82%	0.88	(0.71, 1)	0.09	(0, 0.16)
	Xu et al. (2013) - Ages 7-9	44%	0.95	(0.82, 1)	0.05	(0, 0.12)
	Xu et al. (2013) - Ages 10-12	32%	0.93	(0.78, 1)	0.06	(0, 0.12)
	Xu et al. (2013) - Ages 13-15	74%	0.85	(0.70, 0.97)	0.11	(0.05, 0.17)
Nested Factor	Agostino et al. (2010)	100%	0.90	(0.79, 0.98)	0.10	(0.05, 0.16)
	Arán-Filippetti (2013)	73%	0.98	(0.95, 0.99)	0.07	(0.04, 0.1)
	Brydges et al. (2012)	8%	0.94	(0.86, 1)	0.06	(0.01, 0.09)
	Duan et al. (2010)	56%	0.82	(0.70, 0.94)	0.23	(0.13, 0.33)
	Lehto et al. (2003)	77%	0.89	(0.74, 1)	0.09	(0, 0.15)
	Miller et al. (2012)	14%	0.83	(0.74, 0.91)	0.12	(0.09, 0.15)
	Rose et al. (2012)	60%	0.86	(0.74, 0.96)	0.09	(0.05, 0.13)
	Usai et al. (2014)	94%	0.83	(0.63, 0.99)	0.10	(0.03, 0.16)
	Xu et al. (2013) - Ages 7-9	46%	0.95	(0.82, 1)	0.05	(0, 0.11)
	Xu et al. (2013) - Ages 10-12	34%	0.93	(0.80, 1)	0.05	(0, 0.11)
	Xu et al. (2013) - Ages 13-15	76%	0.86	(0.70, 0.98)	0.10	(0.04, 0.16)
Three-Factor	Agostino et al. (2010)	34%	0.96	(0.88, 1)	0.06	(0, 0.13)
	Arán-Filippetti (2013)	65%	0.93	(0.89, 0.96)	0.12	(0.09, 0.15)
	Brydges et al. (2012)	2%	0.93	(0.86, 0.99)	0.06	(0.02, 0.09)
	Duan et al. (2010)	21%	0.87	(0.76, 0.97)	0.22	(0.11, 0.31)

Model	Study	% Converged	CFI	95% CI	RMSEA	95% CI
	Lehto et al. (2003)	86%	0.91	(0.78, 1)	0.07	(0, 0.13)
	Miller et al. (2012)	86%	0.83	(0.74, 0.91)	0.11	(0.08, 0.14)
	Rose et al. (2012)	60%	0.77	(0.63, 0.89)	0.11	(0.07, 0.14)
	Usai et al. (2014)	24%	0.91	(0.75, 1)	0.08	(0, 0.14)
	Xu et al. (2013) - Ages 7-9	15%	0.92	(0.74, 1)	0.06	(0, 0.13)
	Xu et al. (2013) - Ages 10-12	14%	0.93	(0.76, 1)	0.06	(0, 0.12)
	Xu et al. (2013) - Ages 13-15	26%	0.87	(0.72, 1)	0.09	(0, 0.15)
Inhibition-Shifting Merged	Agostino et al. (2010)	47%	0.90	(0.80, 0.98)	0.10	(0.04, 0.15)
	Arán-Filippetti (2013)	100%	0.92	(0.88, 0.95)	0.12	(0.09, 0.15)
	Brydges et al. (2012)	89%	0.90	(0.81, 0.98)	0.07	(0.03, 0.10)
	Duan et al. (2010)	33%	0.81	(0.65, 0.95)	0.23	(0.12, 0.32)
	Lehto et al. (2003)	98%	0.88	(0.72, 1)	0.09	(0, 0.14)
	Miller et al. (2012)	88%	0.78	(0.67, 0.87)	0.13	(0.09, 0.15)
	Rose et al. (2012)	92%	0.73	(0.58, 0.87)	0.11	(0.08, 0.15)
	Usai et al. (2014)	35%	0.84	(0.66, 1)	0.09	(0.01, 0.15)
	Xu et al. (2013) - Ages 7-9	58%	0.90	(0.73, 1)	0.06	(0, 0.12)
	Xu et al. (2013) - Ages 10-12	69%	0.92	(0.75, 1)	0.06	(0, 0.11)
	Xu et al. (2013) - Ages 13-15	98%	0.85	(0.69, 0.98)	0.09	(0.03, 0.14)
Inhibition-Updating Merged	Agostino et al. (2010)	59%	0.90	(0.80, 0.98)	0.10	(0.04, 0.15)
	Arán-Filippetti (2013)	64%	0.91	(0.87, 0.95)	0.13	(0.10, 0.15)
	Brydges et al. (2012)	6%	0.91	(0.81, 0.98)	0.06	(0.03, 0.09)
	Duan et al. (2010)	35%	0.78	(0.62, 0.92)	0.25	(0.15, 0.35)
	Lehto et al. (2003)	94%	0.87	(0.70, 0.99)	0.09	(0.02, 0.14)
	Miller et al. (2012)	93%	0.74	(0.63, 0.84)	0.13	(0.10, 0.16)
	Rose et al. (2012)	87%	0.74	(0.58, 0.88)	0.11	(0.08, 0.15)
	Usai et al. (2014)	52%	0.84	(0.64, 1)	0.09	(0, 0.15)
	Xu et al. (2013) - Ages 7-9	35%	0.90	(0.71, 1)	0.06	(0, 0.12)
	Xu et al. (2013) - Ages 10-12	40%	0.91	(0.73, 1)	0.06	(0, 0.11)

Model	Study	% Converged	CFI	95% CI	RMSEA	95% CI
	Xu et al. (2013) - Ages 13-15	78%	0.81	(0.65, 0.95)	0.11	(0.05, 0.15)
Shifting-Updating Merged	Agostino et al. (2010)	99%	0.96	(0.87, 1)	0.06	(0, 0.12)
	Arán-Filippetti (2013)	100%	0.92	(0.88, 0.95)	0.12	(0.09, 0.14)
	Brydges et al. (2012)	54%	0.90	(0.80, 0.98)	0.07	(0.03, 0.10)
	Duan et al. (2010)	44%	0.78	(0.64, 0.92)	0.24	(0.15, 0.33)
	Lehto et al. (2003)	97%	0.87	(0.72, 0.99)	0.09	(0.02, 0.14)
	Miller et al. (2012)	98%	0.80	(0.70, 0.89)	0.12	(0.09, 0.15)
	Rose et al. (2012)	90%	0.72	(0.56, 0.85)	0.12	(0.08, 0.15)
	Usai et al. (2014)	75%	0.89	(0.72, 1)	0.07	(0, 0.14)
	Xu et al. (2013) - Ages 7-9	56%	0.90	(0.71, 1)	0.06	(0, 0.12)
	Xu et al. (2013) - Ages 10-12	51%	0.91	(0.75, 1)	0.06	(0, 0.11)
	Xu et al. (2013) - Ages 13-15	51%	0.81	(0.63, 0.95)	0.11	(0.05, 0.16)
Unidimensional	Agostino et al. (2010)	100%	0.90	(0.79, 0.98)	0.09	(0.04, 0.14)
	Arán-Filippetti (2013)	100%	0.91	(0.87, 0.94)	0.12	(0.10, 0.15)
	Brydges et al. (2012)	100%	0.90	(0.80, 0.98)	0.07	(0.03, 0.1)
	Duan et al. (2010)	74%	0.73	(0.56, 0.89)	0.26	(0.16, 0.35)
	Lehto et al. (2003)	100%	0.83	(0.66, 0.97)	0.10	(0.04, 0.15)
	Miller et al. (2012)	99%	0.73	(0.61, 0.83)	0.14	(0.11, 0.17)
	Rose et al. (2012)	100%	0.70	(0.53, 0.84)	0.12	(0.08, 0.15)
	Usai et al. (2014)	95%	0.83	(0.64, 1)	0.09	(0, 0.14)
	Xu et al. (2013) - Ages 7-9	100%	0.90	(0.72, 1)	0.06	(0, 0.12)
	Xu et al. (2013) - Ages 10-12	100%	0.91	(0.74, 1)	0.06	(0, 0.11)
	Xu et al. (2013) - Ages 13-15	100%	0.80	(0.63, 0.95)	0.10	(0.05, 0.15)

Note. CFI = Comparative Fit Index; CI = Confidence Interval; RMSEA = Root Mean Square Error of Approximation.

Table 9. *Adult Studies: Mean Fit Indices (95% CIs) for Converged Models by Measurement Model and Study*

Model	Study	% Converged	CFI	95% CI	RMSEA	95% CI
Bifactor	de Frias et al. (2009) - CE	68%	0.89	(0.65, 1)	0.09	(0, 0.2)
	de Frias et al. (2009) - CN	98%	0.95	(0.86, 1)	0.05	(0, 0.11)
	Fleming et al. (2016)	8%	0.94	(0.9, 0.98)	0.05	(0.03, 0.08)
	Friedman et al. (2011)	13%	0.96	(0.94, 0.99)	0.05	(0.03, 0.07)
	Hedden & Yoon (2006)	5%	0.88	(0.73, 1)	0.09	(0, 0.14)
	Hull et al. (2008)	6%	0.74	(0.63, 0.85)	0.15	(0.1, 0.19)
	Ito et al. (2015)	6%	0.95	(0.91, 0.99)	0.06	(0.03, 0.08)
	Klauer et al (2010) - Study 1	21%	0.87	(0.69, 1)	0.08	(0, 0.14)
	Klauer et al (2010) - Study 2	5%	0.90	(0.79, 0.99)	0.08	(0.03, 0.12)
	Miyake et al. (2000)	13%	0.90	(0.76, 1)	0.07	(0, 0.11)
Nested Factor	de Frias et al. (2009) - CE	71%	0.9	(0.65, 1)	0.08	(0, 0.18)
	de Frias et al. (2009) - CN	98%	0.96	(0.86, 1)	0.05	(0, 0.1)
	Fleming et al. (2016)	89%	0.93	(0.86, 0.98)	0.06	(0.03, 0.08)
	Friedman et al. (2011)	22%	0.91	(0.87, 0.94)	0.08	(0.06, 0.09)
	Hedden & Yoon (2006)	5%	0.87	(0.67, 1)	0.09	(0.01, 0.14)
	Hull et al. (2008)	25%	0.71	(0.59, 0.84)	0.15	(0.11, 0.19)
	Ito et al. (2015)	96%	0.93	(0.88, 0.97)	0.07	(0.05, 0.09)
	Klauer et al (2010) - Study 1	22%	0.88	(0.67, 1)	0.07	(0, 0.14)
	Klauer et al (2010) - Study 2	34%	0.83	(0.72, 0.93)	0.1	(0.06, 0.14)
	Miyake et al. (2000)	34%	0.86	(0.69, 1)	0.07	(0, 0.12)
Three-Factor	de Frias et al. (2009) - CE	7%	0.93	(0.71, 1)	0.07	(0, 0.18)
	de Frias et al. (2009) - CN	37%	0.96	(0.88, 1)	0.05	(0, 0.1)
	Fleming et al. (2016)	94%	0.9	(0.83, 0.96)	0.06	(0.04, 0.08)
	Friedman et al. (2011)	82%	0.94	(0.91, 0.97)	0.06	(0.04, 0.07)
	Hedden & Yoon (2006)	29%	0.81	(0.64, 0.97)	0.1	(0.04, 0.14)
	Hull et al. (2008)	17%	0.62	(0.48, 0.76)	0.16	(0.13, 0.2)
	Ito et al. (2015)	52%	0.92	(0.87, 0.96)	0.06	(0.05, 0.08)

	Klauer et al (2010) - Study 1	12%	0.84	(0.63, 1)	0.08	(0, 0.14)
	Klauer et al (2010) - Study 2	43%	0.82	(0.71, 0.93)	0.1	(0.06, 0.14)
	Miyake et al. (2000)	83%	0.84	(0.67, 0.99)	0.07	(0.01, 0.11)
Inhibition-Shifting Merged	de Frias et al. (2009) - CE	43%	0.91	(0.69, 1)	0.06	(0, 0.16)
	de Frias et al. (2009) - CN	91%	0.94	(0.84, 1)	0.05	(0, 0.1)
	Fleming et al. (2016)	98%	0.79	(0.7, 0.87)	0.09	(0.07, 0.11)
	Friedman et al. (2011)	100%	0.89	(0.84, 0.92)	0.08	(0.06, 0.09)
	Hedden & Yoon (2006)	67%	0.79	(0.57, 0.96)	0.1	(0.04, 0.14)
	Hull et al. (2008)	60%	0.56	(0.42, 0.71)	0.17	(0.13, 0.2)
	Ito et al. (2015)	100%	0.78	(0.71, 0.84)	0.11	(0.09, 0.12)
	Klauer et al (2010) - Study 1	27%	0.67	(0.43, 0.9)	0.11	(0.05, 0.16)
	Klauer et al (2010) - Study 2	80%	0.76	(0.64, 0.88)	0.11	(0.08, 0.15)
	Miyake et al. (2000)	94%	0.74	(0.56, 0.92)	0.09	(0.05, 0.13)
Inhibition-Updating Merged	de Frias et al. (2009) - CE	33%	0.86	(0.59, 1)	0.09	(0, 0.19)
	de Frias et al. (2009) - CN	94%	0.94	(0.84, 1)	0.05	(0, 0.1)
	Fleming et al. (2016)	100%	0.84	(0.75, 0.91)	0.08	(0.06, 0.1)
	Friedman et al. (2011)	77%	0.75	(0.69, 0.81)	0.12	(0.1, 0.13)
	Hedden & Yoon (2006)	60%	0.79	(0.57, 0.95)	0.1	(0.04, 0.14)
	Hull et al. (2008)	24%	0.57	(0.43, 0.73)	0.17	(0.13, 0.2)
	Ito et al. (2015)	100%	0.89	(0.83, 0.94)	0.07	(0.05, 0.09)
	Klauer et al (2010) - Study 1	44%	0.76	(0.55, 0.97)	0.09	(0.03, 0.15)
	Klauer et al (2010) - Study 2	56%	0.71	(0.57, 0.85)	0.12	(0.09, 0.16)
	Miyake et al. (2000)	98%	0.8	(0.61, 0.97)	0.08	(0.03, 0.12)
Shifting-Updating Merged	de Frias et al. (2009) - CE	35%	0.86	(0.61, 1)	0.09	(0, 0.18)
	de Frias et al. (2009) - CN	50%	0.88	(0.74, 0.99)	0.08	(0.03, 0.13)
	Fleming et al. (2016)	97%	0.57	(0.47, 0.67)	0.13	(0.11, 0.15)
	Friedman et al. (2011)	100%	0.89	(0.85, 0.92)	0.08	(0.06, 0.09)
	Hedden & Yoon (2006)	62%	0.79	(0.6, 0.96)	0.09	(0.04, 0.14)
	Hull et al. (2008)	54%	0.48	(0.34, 0.64)	0.18	(0.15, 0.22)

	Ito et al. (2015)	69%	0.6	(0.51, 0.68)	0.14	(0.12, 0.16)
	Klauer et al (2010) - Study 1	40%	0.64	(0.38, 0.86)	0.12	(0.07, 0.17)
	Klauer et al (2010) - Study 2	75%	0.73	(0.6, 0.85)	0.12	(0.08, 0.15)
	Miyake et al. (2000)	93%	0.75	(0.55, 0.92)	0.09	(0.05, 0.13)
Unidimensional	de Frias et al. (2009) - CE	76%	0.85	(0.57, 1)	0.09	(0, 0.18)
	de Frias et al. (2009) - CN	98%	0.88	(0.75, 0.99)	0.08	(0.02, 0.12)
	Fleming et al. (2016)	100%	0.55	(0.45, 0.66)	0.13	(0.11, 0.15)
	Friedman et al. (2011)	100%	0.75	(0.69, 0.81)	0.11	(0.1, 0.13)
	Hedden & Yoon (2006)	100%	0.78	(0.56, 0.96)	0.09	(0.04, 0.14)
	Hull et al. (2008)	93%	0.45	(0.31, 0.6)	0.19	(0.15, 0.22)
	Ito et al. (2015)	100%	0.6	(0.51, 0.69)	0.14	(0.12, 0.16)
	Klauer et al (2010) - Study 1	81%	0.58	(0.34, 0.82)	0.12	(0.07, 0.17)
	Klauer et al (2010) - Study 2	98%	0.65	(0.5, 0.79)	0.13	(0.1, 0.17)
	Miyake et al. (2000)	100%	0.7	(0.5, 0.88)	0.1	(0.06, 0.13)

Note. CFI = Comparative Fit Index; CI = Confidence Interval; RMSEA = Root Mean Square Error of Approximation.

Table 10. *Child and Adolescent Studies: Inter-factor Correlations and 95% Confidence Intervals for Converged Models*

	Three-Factor						Inh.-Shi. Merged		Inh.-Upd. Merged		Shi.-Upd. Merged	
	Upd. w/ Inh.		Upd. w/ Shi.		Shi. w/ Inh.		Inh.-Shi. w/ Upd.		Inh.-Upd. w/ Shi.		Shi.-Upd. w/ Inh.	
Study	*r*	95% C.I.	*r*	95% CI	*r*	95% CI	*r*	95% CI	*r*	95% CI	*r*	95% CI
Agostino et al. (2010)	0.72	(0.88, 0.91)	0.91	(0.78, 0.99)	0.65	(0.39, 0.86)	0.93	(0.79, 1)	0.92	(0.79, 1)	0.73	(0.55, 0.89)
Arán-Filippetti (2013)	0.89	(0.95, 0.94)	0.94	(0.76, 0.99)	0.9	(0.70, 0.98)	0.94	(0.9, 0.98)	0.96	(0.90, 0.98)	0.90	(0.84, 0.96)
Brydges et al. (2012)	0.64	(0.91, 0.89)	0.89	(0.78, 0.97)	0.83	(0.60, 0.97)	0.89	(0.76, 0.99)	0.96	(0.76, 0.99)	0.84	(0.59, 0.99)
Duan et al. (2010)	0.23	(0.57, 0.78)	0.78	(0.59, 0.94)	0.49	(0.12, 0.79)	0.77	(0.59, 0.93)	0.80	(0.59, 0.93)	0.40	(0.08, 0.74)
Lehto et al. (2003)	0.66	(0.92, 0.38)	0.38	(-0.79, 0.87)	0.36	(-0.78, 0.88)	0.75	(0.49, 0.96)	0.39	(0.49, 0.96)	0.71	(0.45, 0.94)
Miller et al. (2012)	0.42	(0.64, 0.81)	0.81	(0.65, 0.95)	0.53	(0.18, 0.82)	0.75	(0.54, 0.91)	0.83	(0.54, 0.91)	0.45	(0.22, 0.68)
Rose et al. (2012)	0.62	(0.93, 0.61)	0.61	(0.37, 0.85)	0.46	(0.01, 0.79)	0.72	(0.40, 0.97)	0.62	(0.40, 0.97)	0.65	(0.35, 0.95)
Usai et al. (2014)	0.61	(0.89, 0.76)	0.76	(0.50, 0.97)	0.40	(0.07, 0.79)	0.85	(0.58, 0.99)	0.75	(0.58, 0.99)	0.62	(0.36, 0.94)
Xu et al. (2013) - Ages 7-9	0.68	(0.95, 0.80)	0.80	(0.44, 0.98)	0.69	(0.26, 0.97)	0.85	(0.54, 1)	0.86	(0.54, 1)	0.77	(0.43, 0.98)
Xu et al. (2013) - Ages 10-12	0.76	(0.97, 0.74)	0.74	(0.47, 0.97)	0.69	(0.33, 0.96)	0.84	(0.58, 0.99)	0.81	(0.58, 0.99)	0.81	(0.52, 0.99)
Xu et al. (2013) - Ages 13-15	0.71	(0.93, 0.54)	0.54	(0.11, 0.88)	0.79	(0.48, 0.98)	0.71	(0.43, 0.94)	0.77	(0.43, 0.94)	0.85	(0.58, 0.99)

Note. CI = Confidence Interval; Inh. = Inhibition; Shi. = Shifting; Upd. = Updating.

Table 11. *Adult Studies: Inter-factor Correlations and 95% Confidence Intervals for Converged Models*

	Three-Factor						Inh.-Shi. Merged		Inh.-Upd. Merged		Shi.-Upd. Merged	
	Upd. w/ Inh.		Upd. w/ Shi.		Shi. w/ Inh.		Inh.-Shi. w/ Upd.		Inh.-Upd. w/ Shi.		Shi.-Upd. w/ Inh.	
Study	*r*	95% CI	*r*	95% CI	*r*	95% CI	*r*	95% CI	*r*	95% CI	*r*	95% CI
de Frias et al. (2009) - CE	0.48	(-0.22, 0.89)	0.60	(0.26, 0.92)	0.46	(-0.25, 0.89)	0.65	(0.30, 0.97)	0.65	(0.26, 0.97)	0.56	(0.11, 0.96)
de Frias et al. (2009) - CN	0.62	(0.25, 0.93)	0.50	(0.25, 0.74)	0.56	(0.10, 0.93)	0.60	(0.37, 0.87)	0.53	(0.31, 0.79)	0.70	(0.32, 0.98)
Fleming et al. (2016)	0.58	(0.38, 0.80)	0.12	(-0.02, 0.27)	0.48	(0.29, 0.70)	0.21	(0.05, 0.37)	0.21	(0.03, 0.39)	0.60	(0.35, 0.90)
Friedman et al. (2011)	0.38	(0.28, 0.47)	0.76	(0.65, 0.86)	0.78	(0.67, 0.87)	0.48	(0.37, 0.57)	0.93	(0.82, 0.99)	0.54	(0.44, 0.65)
Hedden & Yoon (2006)	0.50	(0.06, 0.91)	0.81	(0.52, 0.98)	0.57	(0.13, 0.91)	0.83	(0.54, 0.99)	0.85	(0.59, 0.99)	0.61	(0.27, 0.96)
Hull et al. (2008)	0.41	(-0.02, 0.80)	-0.22	(-0.49, 0.15)	0.13	(-0.58, 0.72)	-0.10	(-0.53, 0.69)	-0.27	(-0.53, 0.16)	0.47	(0.09, 0.87)
Ito et al. (2015)	0.85	(0.71, 0.95)	0.23	(0.10, 0.35)	0.53	(0.38, 0.68)	0.33	(0.18, 0.49)	0.32	(0.20, 0.45)	0.76	(0.51, 0.99)
Klauer et al (2010) - Study 1	0.60	(0.31, 0.89)	0.21	(-0.17, 0.53)	0.48	(0.09, 0.83)	0.50	(0.11, 0.90)	0.40	(0, 0.78)	0.62	(0.22, 0.97)
Klauer et al (2010) - Study 2	0.34	(-0.02, 0.64)	-0.15	(-0.49, 0.16)	0.37	(-0.17, 0.79)	0.23	(-0.27, 0.62)	-0.15	(-0.47, 0.20)	0.33	(-0.06, 0.70)
Miyake et al. (2000)	0.61	(0.23, 0.93)	0.55	(0.22, 0.85)	0.41	(0.02, 0.78)	0.67	(0.31, 0.97)	0.56	(0.26, 0.86)	0.61	(0.26, 0.92)

Note. CI = Confidence Interval; Inh. = Inhibition; Shi. = Shifting; Upd. = Updating.

Table 12. *Post-hoc Evaluation of Publication Bias: Determining the Rate of Researchers Re-selecting their Originally Accepted Model among 5,000 Bootstrapped Samples*

				Percent of Bootstrapped Samples Meeting Fit Cutoffs				Rate of Model Acceptance based on Fit Thresholds			
				CFI		RMSEA		CFI		RMSEA	
Age Band	Model	Study	% Converged	≥.90	≥.95	≤.08	≤.05	≥.90	≥.95	≤.08	≤.05
Child/Adolescent	Three-Factor	Agostino et al. (2010)	34%	94%	65%	65%	30%	32%	22%	22%	10%
		Arán-Filippetti (2013)	65%	92%	7%	0%	0%	60%	5%	0%	0%
		Duan et al. (2010)	21%	28%	7%	1%	0%	6%	1%	0%	0%
		Lehto et al. (2003)	86%	64%	32%	52%	22%	55%	28%	45%	19%
		Rose et al. (2012)	60%	2%	0%	5%	0%	1%	0%	3%	0%
	Shifting-Updating Merged	Miller et al. (2012)	98%	1%	0%	1%	0%	1%	0%	1%	0%
		Usai et al. (2014)	75%	50%	25%	52%	22%	38%	19%	39%	17%
	Unidimensional	Brydges et al. (2012)	100%	48%	11%	77%	14%	48%	11%	77%	14%
		Xu et al. (2013) - Ages 7-9	100%	55%	31%	67%	32%	55%	31%	67%	32%
		Xu et al. (2013) - Ages 10-12	100%	60%	33%	77%	37%	60%	33%	77%	37%
		Median	81%	53%	18%	52%	18%	43%	15%	31%	12%
		Mean	74%	49%	21%	40%	16%	36%	15%	33%	13%
Adult	Nested Factor	Fleming et al. (2016)	89%	81%	22%	97%	27%	72%	20%	86%	24%
		Friedman et al. (2011)	22%	74%	1%	63%	0%	16%	0%	14%	0%
		Ito et al. (2015)	96%	87%	14%	86%	6%	84%	13%	83%	6%
	Three-Factor	de Frias et al. (2009) - CE	7%	69%	48%	56%	39%	5%	3%	4%	3%
		Klauer et al. (2010) - Study 2	43%	7%	0%	14%	0%	3%	0%	6%	0%
		Miyake et al. (2000)	83%	24%	9%	57%	14%	20%	7%	47%	12%
	Inhibition-Updating Merged	Klauer et al. (2010) - Study 1	44%	12%	4%	30%	8%	5%	2%	13%	4%
	Shifting-Updating Merged	Hedden & Yoon (2006)	62%	13%	4%	28%	5%	8%	2%	17%	3%
		Median	53%	47%	7%	57%	7%	12%	3%	16%	4%
		Mean	56%	46%	13%	54%	12%	27%	6%	34%	7%

Note. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation.

Table 13. *Test-Retest Reliability Estimates for Indicators and Control Variables*

Latent Construct	Test	Test-Retest Reliability**
Inhibition	CWIT: Inhibition	0.71
	CWIT: Inhibition/Switching	0.52
	TWT: Total Achievement Score	0.41
Shifting	TMT: Number-Letter Switching (Completion Time)	0.73
	CS: Free Sorting – Confirmed Correct Sorts	0.51
	DF: Switching – Total Correct	0.22
Fluency	VF: Letter Fluency – Total Correct	0.76
	VF: Category Fluency – Total Correct	0.81
	DF: Filled Dots – Total Correct*	0.62
	DF: Empty Dots Only – Total Correct*	0.73
Criterion Variables	20Q – Total Weighted Achievement	0.19
	WCT – Total Consecutively Correct	0.73
	PVT – Total Achievement Score: Free Inquiry	0.66
Control Variables	**Test**	**Test-Retest Reliability****
Processing Speed	CWIT: Color Naming*	0.86
	CWIT: Word Reading*	0.49
	TMT: Number Sequencing*	0.54
	TMT: Letter Sequencing*	0.48
Vocabulary	WASI Vocabulary	0.88

Note. 20Q = Twenty Questions Test; CWIT = Color-Word Interference Test; CF: Category Fluency; CS = Card Sorting; DF = Design Fluency; LF = Letter Fluency; PVT = Proverb Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency; WASI = Wechsler Abbreviated Scale of Intelligence; WCT = Word Context Test

> * Design Fluency scores for Filled and Empty Dots conditions, the Color Word Interference Test scores for the Color and Word conditions, and the Trail Making Test score for the Number and Letter conditions were combined to form composite scores. The composite scores for each of these tests were used in all analyses.
>
> **Age spans for test-retest calculation: D-KEFS = Ages 20-49; WASI Vocabulary = Ages 17-54

Table 14. *Internal Consistency Values for D-KEFS Scores*

Age Group	TMT: Number + Letter Sequencing	VF: Letter Fluency	VF: Category Fluency	CWIT: Color + Word Naming	CS: Free Sorting – Confirmed Correct Sorts	20Q: Total Weighted Achievement	WCT: Total Consecutively Correct	TWT: Total Achievement	PVT: Total Achievement: Free Inquiry
20-29	0.78	0.85	0.61	0.82	0.78	0.50	0.68	0.62	0.71
30-39	0.78	0.90	0.76	0.75	0.82	0.37	0.67	0.72	0.80
40-49	0.74	0.77	0.63	0.72	0.81	0.33	0.53	0.72	0.76

Note. 20Q = Twenty Questions Test; CS = Card Sorting Test; CWIT = Color-Word Interference Test; PVT = Proverb Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency; WCT = Word Context Test

Table 15. *Descriptive Statistics for Variables included in Measurement and Structural Models*

Indicator	Mean	SD	Skewness	Kurtosis	Min.	Max.
CWIT Interference Time*	0.09	0.96	-0.32	1.10	-3.57	4.04
CWIT Switching Time*	0.00	0.96	-0.84	1.17	-3.25	2.77
TWT Total Achievement	10.26	3.02	-0.09	0.42	1.00	19.00
TMT Switch Time*	0.04	0.95	-1.09	1.82	-3.70	2.71
DF Total Designs - Switch Dots*	0.04	0.98	-0.65	1.22	-3.47	2.72
CS Total Confirmed Sorts	9.96	3.23	-0.60	0.79	1.00	18.00
VF Letters - Total Words*	0.11	1.04	0.41	0.20	-2.52	3.48
VF Category - Total Words*	0.09	1.02	-0.06	0.03	-3.00	2.98
DF Filled + Empty	9.92	3.26	-0.02	-0.61	1.00	19.00
20Q Total Weighted Achievement	10.02	3.09	-0.80	0.32	1.00	16.00
WCT Total Achievement	10.18	2.94	-0.58	0.02	1.00	18.00
PVT Total Achievement	10.08	2.69	-0.75	0.29	1.00	14.00

Note. *Indicates a value that was residualized of variance attributable to a control variable. 20Q = Twenty Questions Test; CS = Card Sorting Test; CWIT = Color-Word Interference Test; PVT = Proverb Test; SD = Standard Deviation; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency; WCT = Word Context Test.

Table 16. *Correlation Matrix for Variables included in Measurement and Structural Models*

	1	2	3	4	5	6	7	8	9	10	11	12
1. CWIT Interference Time^{+}	1											
2. CWIT Switching Time^{+}	$.467^{**}$	1										
3. TWT Total Achievement	$.136^{**}$	$.224^{**}$	1									
4. TMT Switch Time^{+}	$.153^{**}$	$.266^{**}$	.073	1								
5. DF Switch Dots^{+}	$.124^{*}$	$.159^{**}$	$.104^{*}$	.032	1							
6. CS Total Confirmed Sorts	$.121^{*}$	$.195^{**}$	$.223^{**}$	$.184^{**}$	$.250^{**}$	1						
7. VF Letters - Total Correct Words^{+}	.072	-.001	.068	.047	.029	.088	1					
8. VF Category - Total Correct Words^{+}	.026	.094	.029	$.132^{*}$	.039	$.149^{**}$	$.443^{**}$	1				
9. DF Filled + Empty	$.134^{**}$	$.193^{**}$	.069	$.166^{**}$	.032	$.271^{**}$	$.180^{**}$	$.202^{**}$	1			
10. 20Q Total Weighted Achievement	.052	$.170^{**}$	$.131^{*}$	$.169^{**}$	$.137^{**}$	$.226^{**}$	-.031	.102	$.097^{*}$	1		
11. WCT Total Achievement	$.108^{*}$	$.152^{**}$	$.181^{**}$	$.226^{**}$	$.127^{**}$	$.337^{**}$	.060	$.165^{**}$	$.178^{**}$	$.273^{**}$	1	
12. PVT Total Achievement	$.124^{*}$	$.143^{**}$	$.168^{**}$	$.181^{**}$	$.153^{**}$	$.333^{**}$	.103	$.197^{**}$	$.177^{**}$	$.240^{**}$	$.386^{**}$	1

Note. *p < .05; **p < .01; $^{+}$Indicates a value that was residualized of variance attributable to a control variable. 20Q = Twenty Questions Test; CS = Card Sorting Test; CWIT = Color-Word Interference Test; PVT = Proverb Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency; WCT = Word Context Test.

Table 17. *Measurement Model Fit Indices*

Model	χ^2 (*p*)	*df*	AIC	BIC	CFI	TLI	RMSEA (90% C.I.)	Percent Convergence
Three Factors (Inh., Shi., Flu.)	69.017 (.0000)	24	12733.811	12855.374	0.871	0.807	0.066 (0.048-0.085)	99.88%
Three Factors (Inh., Shi., Flu.), VF corr.	44.542 (.0045)	23	12711.336	12836.951	0.938	0.904	0.047 (0.026-0.067)	97.98%
Two Factors (Flu., Inh.=Shi.), VF corr.	70.203 (.0000)	25	12732.997	12850.508	0.871	0.814	0.065 (0.047-0.084)	93.10%
Two Factors (Shi., Inh.=Flu.), VF corr.	76.030 (.0000)	25	12738.825	12856.335	0.854	0.790	0.069 (0.052-0.087)	100%
Two Factors (Inh., Shi.=Flu.), VF corr.	49.804 (.0023)	25	12712.598	12830.108	0.929	0.898	0.048 (0.028-0.068)	100%
Unidimensional (Inh.=Shi.=Flu.), VF corr.	82.683 (.0000)	26	12743.477	12856.935	0.838	0.776	0.072 (0.055-0.089)	100%
Bifactor	25.963 (.1006)	18	12702.757	12848.633	0.977	0.954	0.032 (0.000-0.058)	65.86%
Bifactor without Inh.	Did not converge.							

Note. Flu. = Fluency; Inh. = Inhibition; Shi. = Shifting; VF corr. = Verbal Fluency errors correlated

Table 18. *Structural Model Results*

									Standardized Path Coefficients				
Criterion	Model	*df*	χ^2	AIC	BIC	CFI	TLI	RMSEA (90% C.I.)	Bifactor	Inh.	SS	Flu.	R^2
20Q	Bifactor	26	32.332	14796.58	14954.61	0.984	0.971	0.024 (0.000-0.048)	0.376***				0.141**
	Three paths	29	51.587**	14809.83	14955.71	0.941	0.909	0.043 (0.023-0.062)		-0.123	0.725*	-0.302	0.246
	Path from Inh.	31	71.395***	14825.64	14963.41	0.895	0.847	0.055 (0.039-0.072)		0.237***			0.056*
	Path from Shi.	31	54.405**	14808.65	14946.42	0.939	0.911	0.042 (0.023-0.060)			0.379***		0.143**
	Path from Flu.	Non-Positive Definite Covariance Matrix											
WCT	Bifactor	26	32.741	14748.09	14906.12	0.984	0.972	0.025 (0.000-0.048)	0.492***				0.242***
	Three paths	29	52.880**	14762.23	14908.11	0.943	0.911	0.044 (0.024-0.063)		-0.248	0.915*	-0.254	0.408*
	Path from Inh.	31	99.240***	14804.59	14942.36	0.837	0.763	0.072 (0.056-0.088)		0.309***			0.095*
	Path from Shi.	31	56.602**	14761.95	14899.72	0.939	0.911	0.044 (0.025-0.062)			0.511***		0.261***
	Path from Flu.	Non-Positive Definite Covariance Matrix											
PVT	Bifactor	26	34.580	14651.97	14810	0.979	0.964	0.028 (0.000-0.050)	0.478***				0.229***
	Three paths	29	51.771***	14663.16	14809.03	0.945	0.915	0.043 (0.023-0.062)		-0.187	0.742*	-0.114	0.333**
	Path from Inh.	31	94.468***	14701.86	14839.63	0.846	0.777	0.069 (0.054-0.086)		0.305***			0.093*
	Path from Shi.	31	54.27**	14661.66	14799.43	0.944	0.918	0.042 (0.022-0.060)			0.501***		0.251***
	Path from Flu.	Non-Positive Definite Covariance Matrix											

Note. **p*<.05, ***p*<.01, ****p*<.001; 20Q = Twenty Questions Test; Flu. = Fluency; Inh. = Inhibition; PVT = Proverb Test; Shi. = Shifting; WCT = Word Context Test.

Table 19. *D-KEFS Total Achievement Measures*

Test Name	Condition	Outcome	Test Description
Trail Making Test	Condition 4	Number-Letter Switching	Participants connect a series of dots, switching between alphabetical and numerical sets.
Verbal Fluency Test	Condition 1	Letter Fluency Total Correct	In one minute, participants name as many words as possible that begin with a specific letter.
	Condition 2	Category Fluency Total Correct	In one minute, participants name as many words as possible that fall into a specific category.
	Condition 3	Category Switching Total Correct Responses	In one minute, participants switch between naming as many words as possible from two different categories. This outcome represents the number of words produced.
	Condition 4	Category Switching Total Switching Accuracy	Same as Condition 3, but this outcome represents the number of accurate switches between categories.
Design Fluency Test		Total Correct Composite Scaled Score	Participants complete three trials of making unique designs by connecting a set of dots, with the total designs summed across all trials.
Color-Word Interference Test	Condition 3	Inhibition	Displayed with a series of color words written in an incongruent ink color, participants name the ink color for the full series as quickly as possible.
	Condition 4	Inhibition/Switching	Similar to Condition 3, if the incongruent word is displayed in a small box, the participants read the word, but if no small box is present, they read the ink color.
Sorting Test	Condition 1	Free Sorting Confirmed Correct Sorts	Participants sort cards into undisclosed categories based on their characteristics, with this outcome representing their accuracy.
	Condition 1	Free Sorting Description Score	This outcome represents the accuracy of the participant at describing how s/he sorted the cards.

	Condition 2	Sort Recognition Description Score	When presented with the accurate sorting solution, the participant describes how the cards were sorted.
Twenty Questions Test		Total Weighted Achievement Score	The participant asks a series of questions to identify which picture the examiner pre-selected from a page of images. The fewer the questions, the better; but the outcome weights against early accuracy based on luck.
Word Context Test		Total Consecutively Correct	Participants are provided clues into what a series of neologisms mean. Accuracy is weighted by the number of clues given (fewer lead to higher scores) and the number of items consecutively correct.
Tower Test		Total Achievement Score	The participant moved a series of disks between pegs to match a goal peg position. The number of moves is weighted for each item, with fewer moves resulting in a higher score.
		Move Accuracy Ratio	The total moves by the participant divided by the minimum number of moves to complete the problem.
Proverb Test		Total Achievement Score: Free Inquiry	Participants described the meaning of a series of proverbs, with their accuracy and level of abstraction summed.

Note. D-KEFS = Delis-Kaplan Executive Function System

Table 20. *Base Rates of Low Age-Adjusted D-KEFS Total Achievement Scores for Full Nine-Test Battery in 16-89 year-olds – 16 scores: TMT (1 score), VF (4 scores), DF (1 score), CWIT (2 scores), ST (3 scores), 20Q (1 score), WC (1 score), TWT (2 scores), PT (1 score)*

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*789*	*112*	*131*	*183*	*187*	*36*	*92*	*267*	*217*	*177*
≤25th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	.1	–	–	–	–	2.8	–	–	–	–
13 or more	.8	3.6	–	–	–	8.3	1.1	.7	–	–
12 or more	2.3	10.7	.8	–	–	11.1	2.2	2.2	2.3	.6
11 or more	3.7	16.1	2.3	–	–	11.1	4.3	4.5	3.7	.6
10 or more	6.8	25.9	4.6	1.6	–	22.2	13.0	8.2	4.6	1.1
9 or more	10.8	41.1	8.4	3.3	–	25.0	21.7	13.9	6.0	3.4
8 or more	15.3	51.8	16.8	4.4	–	30.6	34.8	18.0	9.2	5.6
7 or more	20.4	62.5	25.2	7.7	.5	41.7	41.3	24.7	12.9	7.9
6 or more	28.0	72.3	40.5	14.2	1.1	58.3	54.3	34.8	18.4	9.6
5 or more	36.5	84.8	49.6	25.1	6.4	61.1	63.0	44.6	27.2	16.9
4 or more	48.0	90.2	64.9	40.4	15.0	72.2	75.0	56.6	39.6	26.6
3 or more	62.9	96.4	79.4	59.0	30.5	91.7	87.0	70.4	56.2	41.2
2 or more	78.7	98.2	92.4	78.1	52.9	97.2	92.4	85.4	74.2	63.3
1 or more	91.9	100.0	97.7	94.0	79.1	100.0	98.9	94.0	92.2	83.1
No low scores	8.1	–	2.3	6.0	20.9	–	1.1	6.0	7.8	16.9
≤16th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	.1	–	–	–	–	2.8	–	–	–	–

Number of Low Scores	Total Sample	WASI FSIQ ≤89	90-99	100-109	110+	Education ≤8	9-11	12	13-15	16+
N	*789*	*112*	*131*	*183*	*187*	*36*	*92*	*267*	*217*	*177*
13 or more	.3	.9	–	–	–	2.8	–	.4	–	–
12 or more	.8	3.6	–	–	–	2.8	–	1.1	–	–
11 or more	.9	4.5	–	–	–	5.6	2.2	1.1	–	–
10 or more	2.7	11.6	.8	–	–	11.1	5.4	3.7	.9	–
9 or more	5.3	22.3	2.3	.5	–	16.7	8.7	8.2	2.3	.6
8 or more	7.0	30.4	3.1	1.1	–	16.7	15.2	9.0	3.2	2.3
7 or more	10.5	37.5	11.5	2.7	–	25.0	25.0	12.7	5.1	3.4
6 or more	15.1	51.8	18.3	3.8	.5	36.1	31.5	17.6	10.1	4.5
5 or more	20.4	64.3	27.5	7.1	.5	44.4	38.0	25.5	14.7	5.6
4 or more	27.9	73.2	37.4	15.8	3.2	55.6	52.2	33.7	19.4	11.3
3 or more	42.2	87.5	59.5	32.8	7.0	72.2	65.2	50.9	34.1	20.9
2 or more	59.3	92.0	71.0	57.9	24.1	88.9	81.5	67.0	50.2	41.2
1 or more	82.6	99.1	93.9	82.5	59.4	97.2	96.7	88.4	78.3	68.9
No low scores	17.4	0.9	6.1	17.5	40.6	2.8	3.3	11.6	21.7	31.1
≤9th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	.1	–	–	–	–	2.8	–	–	–	–
13 or more	.1	–	–	–	–	2.8	–	–	–	–
12 or more	.1	–	–	–	–	2.8	–	–	–	–
11 or more	.6	2.7	–	–	–	2.8	1.1	1.1	–	–
10 or more	.9	3.6	–	–	–	2.8	2.2	1.1	.5	–
9 or more	1.8	8.0	–	–	–	8.3	5.4	1.5	.9	–
8 or more	3.5	16.1	.8	–	–	13.9	8.7	3.4	2.3	.6
7 or more	4.9	21.4	3.1	.5	–	13.9	10.9	5.6	2.8	1.7
6 or more	7.0	29.5	5.3	.5	–	19.4	13.0	9.7	3.2	1.7

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*789*	*112*	*131*	*183*	*187*	*36*	*92*	*267*	*217*	*177*
5 or more	10.4	42.0	7.6	2.7	–	22.2	23.9	12.7	5.5	3.4
4 or more	16.0	50.9	20.6	4.9	.5	38.9	33.7	19.5	10.1	4.0
3 or more	26.1	67.0	32.1	18.0	1.6	47.2	52.2	31.1	19.4	9.0
2 or more	41.6	83.0	48.9	38.3	9.1	72.2	62.0	45.7	34.1	27.7
1 or more	65.9	92.9	75.6	66.7	37.4	94.4	83.7	72.7	58.5	49.7
No low scores	34.1	7.1	24.4	33.3	62.6	5.6	16.3	27.3	41.5	50.3
≤5th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	–	–	–	–	–	–	–	–	–	–
13 or more	–	–	–	–	–	–	–	–	–	–
12 or more	–	–	–	–	–	–	–	–	–	–
11 or more	–	–	–	–	–	–	–	–	–	–
10 or more	.3	–	–	–	–	2.8	1.1	–	–	–
9 or more	.5	1.8	–	–	–	2.8	1.1	.7	–	–
8 or more	.9	4.5	–	–	–	2.8	3.3	1.1	–	–
7 or more	1.6	7.1	.8	–	–	5.6	5.4	1.9	.5	–
6 or more	2.9	13.4	1.5	–	–	11.1	7.6	2.6	1.8	.6
5 or more	4.8	18.8	3.8	–	–	19.4	13.0	4.9	2.3	.6
4 or more	7.5	26.8	5.3	1.6	–	22.2	20.7	7.5	4.1	1.7
3 or more	14.1	43.8	16.0	6.6	.5	36.1	31.5	16.5	7.8	4.5
2 or more	26.7	66.1	30.5	18.0	4.3	44.4	52.2	31.1	19.8	11.9
1 or more	50.7	83.0	58.8	49.2	23.5	75.0	67.4	56.2	46.1	34.5
No low scores	49.3	17.0	41.2	50.8	76.5	25.0	32.6	43.8	53.9	65.5
≤2nd percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–

Number of Low Scores	Total Sample	WASI FSIQ ≤89	WASI FSIQ 90-99	WASI FSIQ 100-109	WASI FSIQ 110+	Education ≤8	Education 9-11	Education 12	Education 13-15	Education 16+
N	*789*	*112*	*131*	*183*	*187*	*36*	*92*	*267*	*217*	*177*
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	–	–	–	–	–	–	–	–	–	–
13 or more	–	–	–	–	–	–	–	–	–	–
12 or more	–	–	–	–	–	–	–	–	–	–
11 or more	–	–	–	–	–	–	–	–	–	–
10 or more	–	–	–	–	–	–	–	–	–	–
9 or more	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	.3	–	–	–	–	2.8	1.1	–	–	–
6 or more	.8	2.7	.8	–	–	2.8	2.2	.7	.5	–
5 or more	2.3	11.6	.8	–	–	11.1	4.3	3.0	.5	.6
4 or more	4.1	16.1	3.8	.5	–	16.7	9.8	4.5	1.8	.6
3 or more	7.7	25.9	9.2	1.1	–	22.2	19.6	10.1	2.8	1.1
2 or more	16.6	44.6	18.3	8.7	3.2	33.3	32.6	21.0	9.7	6.8
1 or more	36.9	69.6	43.5	27.9	13.4	58.3	58.7	42.7	30.9	19.8
No low scores	63.1	30.4	56.5	72.1	86.6	41.7	41.3	57.3	69.1	80.2

Note. All values represent cumulative percentages except for the rows labeled “No low scores,” which provide the percentage of the normative sample with no scores falling under the low score cutoffs. See Table 19 for list of Total Achievement Scores for each D-KEFS Test. Abbreviations: 20Q = Twenty Questions Test; CWIT = Color-Word Interference Test; DF = Design Fluency Test; PT = Proverb Test; ST= Sorting Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency Test; Word Context Test = WC

Table 21. *Base Rates of Low Age-Adjusted D-KEFS Total Achievement Scores in 16-89 year-olds for the Four-Test Battery – 9 scores: TMT (1 score), VF (4 scores), CWIT (2 scores), TWT (2 scores)*

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*823*	*116*	*135*	*189*	*198*	*39*	*96*	*275*	*225*	*188*
≤25th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	.2	.9	–	–	–	–	–	.4	.4	–
7 or more	1.9	3.4	.7	.5	–	5.1	4.2	2.5	1.3	–
6 or more	5.1	12.1	3.7	2.6	–	15.4	9.4	6.5	2.7	1.6
5 or more	13.0	37.1	15.6	5.3	1.5	20.5	22.9	17.5	8.4	5.3
4 or more	25.5	55.2	28.9	16.9	8.1	41.0	42.7	29.1	20.4	14.4
3 or more	41.6	76.7	46.7	30.7	19.2	64.1	63.5	48.0	34.2	25.0
2 or more	62.2	83.6	71.1	55.0	42.4	87.2	80.2	68.7	52.4	50.0
1 or more	84.9	99.1	89.6	83.6	72.2	97.4	95.8	88.4	81.3	76.1
No low scores	15.1	0.9	10.4	16.4	27.8	2.6	4.2	11.6	18.7	23.9
≤16th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	.7	2.6	–	–	–	2.6	3.1	.7	–	–
6 or more	1.2	2.6	–	.5	–	2.6	3.1	1.1	.9	.5
5 or more	5.2	12.9	5.2	2.6	.5	12.8	8.3	7.3	2.7	2.1
4 or more	12.3	29.3	14.1	5.8	2.0	23.1	24.0	16.0	8.9	2.7
3 or more	23.5	56.9	27.4	11.6	6.1	33.3	45.8	28.0	17.3	10.6
2 or more	43.7	76.7	47.4	35.4	20.2	71.8	61.5	49.5	36.9	28.7
1 or more	71.8	92.2	74.8	69.8	51.0	89.7	87.5	79.3	63.1	59.6
No low scores	28.2	7.8	25.2	30.2	49.0	10.3	12.5	20.7	36.9	40.4
≤9th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*823*	*116*	*135*	*189*	*198*	*39*	*96*	*275*	*225*	*188*
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	.2	–	–	–	–	2.6	1.0	–	–	–
6 or more	.6	1.7	–	–	–	2.6	2.1	.4	.4	–
5 or more	2.1	7.8	.7	1.6	–	5.1	4.2	2.5	.9	1.1
4 or more	6.1	14.7	7.4	2.6	–	10.3	11.5	7.6	4.4	2.1
3 or more	13.6	34.5	15.6	7.4	1.5	28.2	27.1	15.3	10.2	5.3
2 or more	28.2	61.2	29.6	23.8	6.1	51.3	47.9	30.2	22.7	17.0
1 or more	55.0	85.3	54.8	50.3	33.3	84.6	71.9	58.5	48.4	43.1
No low scores	45.0	14.7	45.2	49.7	66.7	15.4	28.1	41.5	51.6	56.9
≤5th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	.1	–	–	–	–	–	1.0	–	–	–
6 or more	.2	–	–	–	–	2.6	1.0	–	–	–
5 or more	.7	2.6	–	–	–	2.6	2.1	.7	.4	–
4 or more	2.3	6.9	2.2	.5	–	2.6	6.3	2.2	1.8	1.1
3 or more	6.0	16.4	6.7	1.6	–	15.4	15.6	6.5	3.1	1.6
2 or more	15.9	37.1	17.0	9.0	2.0	33.3	31.3	16.4	12.4	8.0
1 or more	42.0	76.7	44.4	33.3	20.7	69.2	59.4	44.7	37.3	29.3
No low scores	58.0	23.3	55.6	66.7	79.3	30.8	40.6	55.3	62.7	70.7
≤2nd percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	–	–	–	–	–	–	–	–	–	–
6 or more	.1	–	–	–	–	–	1.0	–	–	–
5 or more	.1	–	–	–	–	–	1.0	–	–	–

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*823*	*116*	*135*	*189*	*198*	*39*	*96*	*275*	*225*	*188*
4 or more	.5	1.7	–	–	–	2.6	1.0	.7	–	–
3 or more	3.3	9.5	4.4	–	–	7.7	7.3	4.7	1.8	–
2 or more	9.2	23.3	11.9	3.7	.5	15.4	17.7	10.2	6.7	5.3
1 or more	30.1	56.9	37.8	18.5	10.6	48.7	49.0	33.1	26.7	16.5
No low scores	69.9	43.1	62.2	81.5	89.4	51.3	51.0	66.9	73.3	83.5

Note. All values represent cumulative percentages except for the rows labeled “No low scores,” which provide the percentage of the normative sample with no scores falling under the low score cutoffs. See Table 19 for list of Total Achievement Scores for each D-KEFS Test. Abbreviations: CWIT = Color-Word Interference Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency Test

Table 22. *Base Rates of Low Age-Adjusted D-KEFS Total Achievement Scores in 16-89 year-olds for the Three-Test Battery – 7 scores: TMT (1 score), VF (4 scores), CWIT (2 scores)*

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*1028*	*160*	*163*	*243*	*241*	*57*	*125*	*351*	*275*	*220*
≤25th percentile										
7 low scores	.3	1.3	–	–	–	3.5	–	–	.4	–
6 or more	2.9	7.5	1.8	1.2	.4	10.5	5.6	3.4	.7	1.4
5 or more	8.4	22.5	6.1	5.3	.8	21.1	17.6	8.5	4.7	4.1
4 or more	18.2	44.4	18.4	11.9	2.9	33.3	36.8	19.9	12.4	8.2
3 or more	32.3	63.8	33.7	25.9	10.8	52.6	50.4	35.6	25.5	20.0
2 or more	50.9	83.1	58.3	42.0	27.8	77.2	70.4	57.8	42.2	32.7
1 or more	76.6	95.0	85.9	71.2	60.2	96.5	92.0	81.8	70.5	61.8
No low scores	23.4	5.0	14.1	28.8	39.8	3.5	8.0	18.2	29.5	38.2
≤16th percentile										
7 low scores	.2	1.3	–	–	–	3.5	–	–	–	–
6 or more	1.5	4.4	1.2	.4	–	7.0	4.8	1.1	–	.5
5 or more	2.8	9.4	1.2	1.6	.4	12.3	5.6	2.6	1.1	1.4
4 or more	8.9	23.1	8.6	4.1	.4	22.8	18.4	11.1	4.7	1.8
3 or more	18.7	44.4	17.8	9.9	2.9	36.8	37.6	20.5	12.7	7.7
2 or more	35.8	67.5	37.4	28.0	14.9	57.9	54.4	38.7	30.2	21.8
1 or more	62.8	88.1	67.5	58.4	38.6	87.7	78.4	70.1	54.9	45.9
No low scores	37.2	11.9	32.5	41.6	61.4	12.3	21.6	29.9	45.1	54.1
≤9th percentile										
7 low scores	–	–	–	–	–	–	–	–	–	–
6 or more	.5	1.3	–	–	–	5.3	1.6	–	–	–
5 or more	1.6	5.6	.6	.8	–	7.0	3.2	1.4	.4	.9
4 or more	5.2	15.0	4.9	2.1	–	15.8	9.6	6.6	2.2	1.4
3 or more	11.4	28.8	12.3	5.3	1.2	26.3	24.0	13.1	6.9	3.2

Number of Low Scores	Total Sample	WASI FSIQ				Education				
		≤89	90-99	100-109	110+	≤8	9-11	12	13-15	16+
N	*1028*	*160*	*163*	*243*	*241*	*57*	*125*	*351*	*275*	*220*
2 or more	24.4	51.9	25.2	18.5	6.6	49.1	42.4	24.8	19.6	13.2
1 or more	48.1	78.1	51.5	41.6	23.7	77.2	64.8	51.0	41.1	35.0
No low scores	51.9	21.9	48.5	58.4	76.3	22.8	35.2	49.0	58.9	65.0
≤5th percentile										
7 low scores	–	–	–	–	–	–	–	–	–	
6 or more	.1	–	–	–	–	–	.8	–	–	
5 or more	.3	–	–	–	–	1.8	1.6	–	–	
4 or more	1.9	6.3	1.2	.8	–	5.3	4.8	2.6	.4	.5
3 or more	6.0	16.3	7.4	2.5	–	17.5	14.4	6.6	2.5	1.8
2 or more	14.0	33.8	13.5	7.8	2.5	33.3	28.0	13.7	9.5	7.3
1 or more	36.1	66.9	41.1	25.5	16.2	59.6	53.6	39.9	29.5	22.3
No low scores	63.9	33.1	58.9	74.5	83.8	40.4	46.4	60.1	70.5	77.7
≤2nd percentile										
7 low scores	–	–	–	–	–	–	–	–	–	–
6 or more	–	–	–	–	–	–	–	–	–	–
5 or more	.1	–	–	–	–	–	.8	–	–	–
4 or more	.5	1.3	–	.4	–	–	2.4	.6	–	–
3 or more	2.6	6.9	4.3	.4	–	10.5	5.6	3.7	.4	–
2 or more	8.9	23.1	10.4	3.7	1.2	19.3	16.0	10.3	4.7	5.0
1 or more	26.8	50.6	33.1	16.5	9.1	42.1	43.2	30.2	21.1	15.5
No low scores	73.2	49.4	66.9	83.5	90.9	57.9	56.8	69.8	78.9	84.5

Note. All values represent cumulative percentages except for the rows labeled “No low scores,” which provide the percentage of the normative sample with no scores falling under the low score cutoffs. See Table 19 for list of Total Achievement Scores for each D-KEFS Test. Abbreviations: CWIT = Color-Word Interference Test; TMT = Trail Making Test; VF = Verbal Fluency Test

Table 23. *Base Rates of Low Age-Adjusted D-KEFS Total Achievement Scores in 16-69 year-olds for Full Nine-Test Battery – 16 scores: TMT (1 score), VF (4 scores), DF (1 score), CWIT (2 scores), ST (3 scores), 20Q (1 score), WC (1 score), TWT (2 scores), PT (1 score)*

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*638*	*100*	*101*	*146*	*140*	*18*	*72*	*209*	*189*	*150*
≤25th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	–	–	–	–	–	–	–	–	–	–
13 or more	.3	2.0	–	–	–	5.6	–	.5	–	–
12 or more	2.0	10.0	1.0	–	–	5.6	1.4	2.4	2.6	.7
11 or more	3.4	16.0	2.0	–	–	5.6	4.2	4.8	3.7	.7
10 or more	6.4	25.0	2.0	1.4	–	22.2	13.9	8.6	4.2	.7
9 or more	10.2	39.0	3.0	3.4	–	22.2	20.8	14.4	5.8	3.3
8 or more	15.0	49.0	12.9	4.1	–	33.3	34.7	19.1	9.5	4.7
7 or more	20.7	60.0	21.8	7.5	.7	44.4	43.1	27.3	13.8	6.7
6 or more	28.2	70.0	35.6	15.1	1.4	61.1	56.9	36.8	20.1	8.7
5 or more	37.3	84.0	45.5	27.4	5.7	61.1	65.3	47.4	29.1	17.3
4 or more	49.2	90.0	61.4	43.8	13.6	72.2	75.0	59.3	42.9	28.0
3 or more	63.0	96.0	77.2	60.3	27.1	94.4	86.1	72.7	58.7	40.0
2 or more	79.3	98.0	93.1	78.1	52.9	94.4	91.7	88.5	76.2	62.7
1 or more	91.7	100.0	99.0	93.8	75.7	100.0	98.6	94.7	92.1	82.7
No low scores	8.3	–	1.0	6.2	24.3	–	1.4	5.3	7.9	17.3
≤16th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	–	–	–	–	–	–	–	–	–	–
13 or more	–	–	–	–	–	–	–	–	–	–

Number of Low Scores	Total Sample	WASI FSIQ ≤89	90-99	100-109	110+	Education ≤8	9-11	12	13-15	16+
N	*638*	*100*	*101*	*146*	*140*	*18*	*72*	*209*	*189*	*150*
12 or more	.5	3.0	–	–	–	–	1.4	1.0	–	–
11 or more	.5	3.0	–	–	–	–	1.4	1.0	–	–
10 or more	2.4	11.0	–	–	–	11.1	5.6	3.8	.5	–
9 or more	5.2	22.0	1.0	.7	–	16.7	8.3	9.1	2.1	.7
8 or more	6.7	28.0	2.0	1.4	–	16.7	13.9	10.0	3.2	2.0
7 or more	10.0	35.0	7.9	2.7	–	27.8	23.6	13.4	4.8	3.3
6 or more	15.0	50.0	14.9	4.1	.7	33.3	30.6	19.6	10.6	4.7
5 or more	20.2	62.0	21.8	7.5	.7	38.9	37.5	27.3	15.9	5.3
4 or more	28.5	72.0	32.7	17.1	3.6	55.6	54.2	36.8	20.6	11.3
3 or more	43.9	87.0	57.4	35.6	6.4	72.2	70.8	54.5	37.0	21.3
2 or more	61.1	92.0	70.3	60.3	24.3	94.4	83.3	70.8	52.9	43.3
1 or more	83.1	99.0	95.0	82.9	56.4	100.0	97.2	90.4	79.4	68.7
No low scores	16.9	1.0	5.0	17.1	43.6	–	2.8	9.6	20.6	31.3
≤9th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	–	–	–	–	–	–	–	–	–	–
13 or more	–	–	–	–	–	–	–	–	–	–
12 or more	–	–	–	–	–	–	–	–	–	–
11 or more	.3	2.0	–	–	–	–	–	1.0	–	–
10 or more	.5	3.0	–	–	–	–	1.4	1.0	–	–
9 or more	1.3	7.0	–	–	–	–	5.6	1.4	.5	–
8 or more	3.3	16.0	–	–	–	11.1	9.7	3.3	2.1	.7
7 or more	4.7	20.0	3.0	.7	–	11.1	9.7	6.2	2.6	2.0
6 or more	6.9	28.0	5.0	.7	–	22.2	11.1	11.0	3.2	2.0
5 or more	10.2	41.0	5.0	2.1	–	27.8	23.6	13.9	5.3	2.7

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*638*	*100*	*101*	*146*	*140*	*18*	*72*	*209*	*189*	*150*
4 or more	15.5	48.0	16.8	4.8	–	44.4	33.3	20.6	10.1	3.3
3 or more	26.2	66.0	25.7	18.5	1.4	50.0	55.6	32.5	20.1	8.0
2 or more	43.6	83.0	46.5	40.4	10.7	77.8	63.9	50.2	36.5	29.3
1 or more	66.6	93.0	74.3	67.1	37.1	100.0	83.3	74.6	60.8	50.7
No low scores	33.4	7.0	25.7	32.9	62.9	–	16.7	25.4	39.2	49.3
≤5th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	–	–	–	–	–	–	–	–	–	–
13 or more	–	–	–	–	–	–	–	–	–	–
12 or more	–	–	–	–	–	–	–	–	–	–
11 or more	–	–	–	–	–	–	–	–	–	–
10 or more	–	–	–	–	–	–	–	–	–	–
9 or more	.2	1.0	–	–	–	–	–	.5	–	–
8 or more	.6	4.0	–	–	–	–	2.8	1.0	–	–
7 or more	1.4	6.0	1.0	–	–	–	5.6	1.9	.5	–
6 or more	2.5	12.0	2.0	–	–	5.6	6.9	2.9	1.6	.7
5 or more	4.4	17.0	3.0	–	–	22.2	11.1	5.3	2.1	.7
4 or more	7.2	25.0	4.0	1.4	–	27.8	19.4	8.1	4.2	1.3
3 or more	13.8	41.0	12.9	6.8	.7	38.9	31.9	17.2	8.5	4.0
2 or more	27.4	65.0	25.7	19.2	5.0	50.0	54.2	33.5	20.6	12.0
1 or more	52.4	83.0	58.4	51.4	23.6	83.3	69.4	58.9	48.7	36.0
No low scores	47.6	17.0	41.6	48.6	76.4	16.7	30.6	41.1	51.3	64.0
≤2nd percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–

Number of Low Scores	Total Sample	WASI FSIQ ≤89	WASI FSIQ 90-99	WASI FSIQ 100-109	WASI FSIQ 110+	Education ≤8	Education 9-11	Education 12	Education 13-15	Education 16+
N	*638*	*100*	*101*	*146*	*140*	*18*	*72*	*209*	*189*	*150*
14 or more	–	–	–	–	–	–	–	–	–	–
13 or more	–	–	–	–	–	–	–	–	–	–
12 or more	–	–	–	–	–	–	–	–	–	–
11 or more	–	–	–	–	–	–	–	–	–	–
10 or more	–	–	–	–	–	–	–	–	–	–
9 or more	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	–	–	–	–	–	–	–	–	–	–
6 or more	.6	3.0	–	–	–	–	1.4	1.0	.5	–
5 or more	1.9	10.0	1.0	–	–	5.6	2.8	3.3	.5	.7
4 or more	3.6	15.0	3.0	.7	–	16.7	9.7	4.3	1.6	.7
3 or more	7.4	24.0	6.9	1.4	–	27.8	19.4	10.5	2.6	.7
2 or more	17.1	44.0	14.9	9.6	4.3	44.4	34.7	23.0	10.1	6.0
1 or more	37.6	68.0	42.6	27.4	13.6	55.6	62.5	45.9	31.2	20.0
No low scores	62.4	32.0	57.4	72.6	86.4	44.4	37.5	54.1	68.8	80.0

Note. All values represent cumulative percentages except for the rows labeled "No low scores," which provide the percentage of the normative sample with no scores falling under the low score cutoffs. See Table 19 for list of Total Achievement Scores for each D-KEFS Test. Abbreviations: 20Q = Twenty Questions Test; CWIT = Color-Word Interference Test; DF = Design Fluency Test; PT = Proverb Test; ST= Sorting Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency Test; Word Context Test = WC

Table 24. *Base Rates of Low Age-Adjusted D-KEFS Total Achievement Scores in 16-69 year-olds for the Four-Test Battery – 9 scores: TMT (1 score), VF (4 scores), CWIT (2 scores), TWT (2 scores)*

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*667*	*103*	*104*	*151*	*149*	*21*	*75*	*215*	*196*	*160*
≤25th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	.3	1.0	–	–	–	–	–	.5	.5	–
7 or more	1.9	2.9	1.0	.7	–	4.8	4.0	2.8	1.5	–
6 or more	5.1	10.7	2.9	3.3	–	14.3	9.3	7.0	3.1	1.9
5 or more	12.7	36.9	12.5	6.0	.7	19.0	24.0	18.6	8.7	3.8
4 or more	25.0	55.3	24.0	17.2	7.4	38.1	44.0	29.8	20.4	13.8
3 or more	41.1	75.7	42.3	30.5	17.4	66.7	61.3	48.8	36.2	23.8
2 or more	61.9	82.5	68.3	55.0	41.6	85.7	76.0	70.2	54.6	50.0
1 or more	84.6	99.0	89.4	82.8	69.8	95.2	94.7	88.8	82.7	75.0
No low scores	15.4	1.0	10.6	17.2	30.2	4.8	5.3	11.2	17.3	25.0
≤16th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	.4	–	–	–	–	–	2.7	.5	–	–
6 or more	1.0	1.9	–	.7	–	–	2.7	.9	1.0	.6
5 or more	4.8	12.6	2.9	2.6	.7	9.5	9.3	7.4	2.6	1.3
4 or more	11.8	29.1	11.5	5.3	1.3	23.8	22.7	17.2	8.7	1.9
3 or more	22.9	55.3	23.1	10.6	5.4	33.3	45.3	28.8	17.9	9.4
2 or more	44.8	75.7	46.2	36.4	20.8	71.4	62.7	53.0	38.8	29.4
1 or more	71.4	92.2	72.1	69.5	48.3	95.2	86.7	80.0	63.3	59.4

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*667*	*103*	*104*	*151*	*149*	*21*	*75*	*215*	*196*	*160*
No low scores	28.6	7.8	27.9	30.5	51.7	4.8	13.3	20.0	36.7	40.6
≤9th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	–	–	–	–	–	–	–	–	–	–
6 or more	.4	1.9	–	–	–	–	1.3	.5	.5	–
5 or more	1.6	7.8	–	1.3	–	4.8	4.0	2.3	.5	.6
4 or more	5.4	13.6	4.8	2.0	–	4.8	10.7	7.9	4.1	1.3
3 or more	12.9	33.0	12.5	5.3	.7	33.3	28.0	15.3	9.7	3.8
2 or more	28.8	60.2	26.0	24.5	6.7	57.1	48.0	33.0	23.5	16.9
1 or more	55.3	85.4	52.9	49.7	32.2	90.5	70.7	60.9	50.0	42.5
No low scores	44.7	14.6	47.1	50.3	67.8	9.5	29.3	39.1	50.0	57.5
≤5th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	–	–	–	–	–	–	–	–	–	–
6 or more	–	–	–	–	–	–	–	–	–	–
5 or more	.4	1.9	–	–	–	–	1.3	.5	.5	–
4 or more	1.8	6.8	1.0	–	–	–	5.3	2.3	1.5	–
3 or more	5.4	15.5	3.8	1.3	–	14.3	16.0	6.5	3.1	.6
2 or more	15.4	35.0	13.5	7.9	2.0	38.1	30.7	16.3	12.8	7.5
1 or more	42.7	76.7	41.3	33.8	20.1	71.4	58.7	47.4	39.3	29.4
No low scores	57.3	23.3	58.7	66.2	79.9	28.6	41.3	52.6	60.7	70.6
≤2nd percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–

Number of Low Scores	Total Sample	WASI FSIQ ≤89	WASI FSIQ 90-99	WASI FSIQ 100-109	WASI FSIQ 110+	Education ≤8	Education 9-11	Education 12	Education 13-15	Education 16+
N	*667*	*103*	*104*	*151*	*149*	*21*	*75*	*215*	*196*	*160*
7 or more	–	–	–	–	–	–	–	–	–	–
6 or more	–	–	–	–	–	–	–	–	–	–
5 or more	–	–	–	–	–	–	–	–	–	–
4 or more	.3	1.9	–	–	–	–	–	.9	–	–
3 or more	2.8	8.7	2.9	–	–	4.8	8.0	4.2	1.5	–
2 or more	9.0	23.3	9.6	3.3	.7	14.3	18.7	10.7	6.6	4.4
1 or more	30.4	54.4	35.6	17.9	10.7	42.9	49.3	35.8	28.1	15.6
No low scores	69.6	45.6	64.4	82.1	89.3	57.1	50.7	64.2	71.9	84.4

Note. All values represent cumulative percentages except for the rows labeled "No low scores," which provide the percentage of the normative sample with no scores falling under the low score cutoffs. See Table 19 for list of Total Achievement Scores for each D-KEFS Test. Abbreviations: CWIT = Color-Word Interference Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency Test

Table 25. *Base Rates of Low Age-Adjusted D-KEFS Total Achievement Scores in 16-69 year-olds for the Three-Test Battery – 7 scores: TMT (1 score), VF (4 scores), CWIT (2 scores)*

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*814*	*133*	*121*	*191*	*179*	*30*	*90*	*270*	*236*	*188*
≤25th percentile										
7 low scores	.2	.8	–	–	–	3.3	–	–	.4	–
6 or more	2.9	6.8	1.7	1.6	.6	6.7	6.7	4.1	.8	1.6
5 or more	8.1	21.8	5.0	5.2	1.1	20.0	18.9	9.3	4.7	3.7
4 or more	17.3	40.6	14.9	12.6	2.2	33.3	35.6	20.7	12.7	6.9
3 or more	31.8	62.4	29.8	26.2	10.1	56.7	51.1	36.7	26.3	18.6
2 or more	50.5	82.0	56.2	42.4	26.3	80.0	68.9	60.4	43.6	31.4
1 or more	75.3	94.7	85.1	69.1	56.4	93.3	88.9	82.6	72.9	58.5
No low scores	24.7	5.3	14.9	30.9	43.6	6.7	11.1	17.4	27.1	41.5
≤16th percentile										
7 low scores	.1	.8	–	–	–	3.3	–	–	–	–
6 or more	1.2	3.8	.8	.5	–	3.3	5.6	1.1	–	.5
5 or more	2.6	9.0	.8	1.0	.6	10.0	5.6	3.0	1.3	1.1
4 or more	8.5	23.3	5.8	4.2	.6	26.7	18.9	11.9	4.2	1.1
3 or more	18.1	40.6	15.7	9.9	2.2	40.0	37.8	21.9	12.3	6.9
2 or more	36.1	65.4	36.4	28.3	15.1	66.7	55.6	40.7	31.4	21.3
1 or more	62.4	88.0	65.3	57.6	36.9	90.0	77.8	72.2	56.8	43.6
No low scores	37.6	12.0	34.7	42.4	63.1	10.0	22.2	27.8	43.2	56.4
≤9th percentile										
7 low scores	–	–	–	–	–	–	–	–	–	–
6 or more	.2	.8	–	–	–	3.3	1.1	–	–	–

Number of Low Scores	Total Sample	WASI FSIQ ≤89	WASI FSIQ 90-99	WASI FSIQ 100-109	WASI FSIQ 110+	Education ≤8	Education 9-11	Education 12	Education 13-15	Education 16+
N	*814*	*133*	*121*	*191*	*179*	*30*	*90*	*270*	*236*	*188*
5 or more	1.4	6.0	–	.5	–	3.3	3.3	1.9	.4	.5
4 or more	4.7	15.0	2.5	1.6	–	16.7	10.0	6.7	1.7	1.1
3 or more	10.6	25.6	9.1	5.2	1.1	26.7	25.6	13.0	6.4	2.7
2 or more	24.0	48.9	23.1	17.3	7.3	56.7	43.3	25.6	20.3	11.7
1 or more	47.8	75.9	49.6	40.8	22.9	86.7	62.2	53.3	42.8	33.0
No low scores	52.2	24.1	50.4	59.2	77.1	13.3	37.8	46.7	57.2	67.0
≤5th percentile										
7 low scores	–	–	–	–	–	–	–	–	–	–
6 or more	–	–	–	–	–	–	–	–	–	–
5 or more	.1	–	–	–	–	–	1.1	–	–	–
4 or more	1.5	6.0	–	–	–	3.3	4.4	2.2	.4	–
3 or more	5.3	15.8	3.3	2.1	–	20.0	15.6	5.9	2.1	1.1
2 or more	13.0	30.8	10.7	6.8	2.2	40.0	28.9	13.0	9.3	5.9
1 or more	35.5	64.7	37.2	24.6	16.2	63.3	50.0	41.5	31.4	20.7
No low scores	64.5	35.3	62.8	75.4	83.8	36.7	50.0	58.5	68.6	79.3
≤2nd percentile										
7 low scores	–	–	–	–	–	–	–	–	–	–
6 or more	–	–	–	–	–	–	–	–	–	–
5 or more	–	–	–	–	–	–	–	–	–	–
4 or more	.4	1.5	–	–	–	–	1.1	.7	–	–
3 or more	2.2	6.8	2.5	–	–	10.0	5.6	3.3	.4	–
2 or more	8.5	23.3	8.3	3.1	1.7	23.3	17.8	10.0	4.7	4.3
1 or more	26.5	48.1	31.4	15.7	9.5	40.0	43.3	31.9	22.5	13.8
No low scores	73.5	51.9	68.6	84.3	90.5	60.0	56.7	68.1	77.5	86.2

Note. All values represent cumulative percentages except for the rows labeled “No low scores,” which provide the percentage of the normative sample with no scores falling under the low score cutoffs. See Table 19 for list of Total Achievement Scores for each D-KEFS Test. Abbreviations: CWIT = Color-Word Interference Test; TMT = Trail Making Test; VF = Verbal Fluency Test

Table 26. *Base Rates of Low Age-Adjusted D-KEFS Total Achievement Scores in 60-89 year-olds for Full Nine-Test Battery – 16 scores: TMT (1 score), VF (4 scores), DF (1 score), CWIT (2 scores), ST (3 scores), 20Q (1 score), WC (1 score), TWT (2 scores), PT (1 score)*

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*233*	*31*	*41*	*54*	*65*	*21*	*35*	*90*	*43*	*44*
≤25th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	.4	–	–	–	–	4.8	–	–	–	–
13 or more	2.6	12.9	–	–	–	14.3	2.9	2.2	–	–
12 or more	3.4	16.1	–	–	–	19.0	2.9	2.2	–	2.3
11 or more	4.7	19.4	2.4	–	–	19.0	2.9	4.4	2.3	2.3
10 or more	9.4	32.3	9.8	3.7	–	23.8	14.3	8.9	4.7	4.5
9 or more	13.7	45.2	19.5	5.6	–	28.6	25.7	13.3	4.7	6.8
8 or more	16.7	54.8	22.0	7.4	–	28.6	31.4	16.7	4.7	11.4
7 or more	21.0	64.5	31.7	11.1	–	38.1	37.1	21.1	7.0	13.6
6 or more	28.8	80.6	51.2	13.0	–	57.1	48.6	30.0	9.3	15.9
5 or more	36.9	87.1	61.0	22.2	9.2	61.9	60.0	42.2	16.3	15.9
4 or more	46.8	90.3	75.6	33.3	18.5	71.4	74.3	53.3	23.3	22.7
3 or more	62.7	96.8	87.8	55.6	36.9	85.7	88.6	64.4	41.9	47.7
2 or more	76.8	96.8	92.7	75.9	53.8	95.2	91.4	78.9	60.5	68.2
1 or more	92.3	100.0	95.1	94.4	84.6	100.0	100.0	92.2	88.4	86.4
No low scores	7.7	–	4.9	5.6	15.4	–	–	7.8	11.6	13.6
≤16th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–

Number of Low Scores	Total Sample	WASI FSIQ ≤89	90-99	100-109	110+	Education ≤8	9-11	12	13-15	16+
N	*233*	*31*	*41*	*54*	*65*	*21*	*35*	*90*	*43*	*44*
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	.4	–	–	–	–	4.8	–	–	–	–
13 or more	.9	3.2	–	–	–	4.8	–	1.1	–	–
12 or more	1.3	3.2	–	–	–	4.8	2.9	1.1	–	–
11 or more	1.7	6.5	–	–	–	9.5	2.9	1.1	–	–
10 or more	3.9	16.1	2.4	–	–	14.3	2.9	4.4	2.3	–
9 or more	6.4	25.8	4.9	1.9	–	19.0	8.6	6.7	2.3	2.3
8 or more	8.6	38.7	4.9	1.9	–	19.0	17.1	7.8	2.3	4.5
7 or more	13.3	45.2	17.1	5.6	–	23.8	28.6	12.2	4.7	6.8
6 or more	16.3	54.8	24.4	5.6	–	38.1	34.3	13.3	7.0	6.8
5 or more	24.5	74.2	43.9	9.3	–	52.4	45.7	23.3	11.6	9.1
4 or more	29.2	77.4	51.2	16.7	1.5	57.1	48.6	31.1	14.0	11.4
3 or more	40.3	90.3	65.9	29.6	7.7	71.4	54.3	46.7	18.6	22.7
2 or more	56.7	90.3	75.6	51.9	26.2	81.0	74.3	61.1	39.5	38.6
1 or more	82.0	96.8	92.7	79.6	66.2	95.2	97.1	81.1	72.1	75.0
No low scores	18.0	3.2	7.3	20.4	33.8	4.8	2.9	18.9	27.9	25.0
≤9th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	.4	–	–	–	–	4.8	–	–	–	–
13 or more	.4	–	–	–	–	4.8	–	–	–	–
12 or more	.4	–	–	–	–	4.8	–	–	–	–
11 or more	1.3	3.2	–	–	–	4.8	2.9	1.1	–	–
10 or more	1.7	3.2	–	–	–	4.8	2.9	1.1	2.3	–
9 or more	3.4	12.9	–	–	–	14.3	5.7	2.2	2.3	–
8 or more	4.7	19.4	2.4	–	–	19.0	5.7	3.3	2.3	2.3

Number of Low Scores	Total Sample	WASI FSIQ ≤89	90-99	100-109	110+	Education ≤8	9-11	12	13-15	16+
N	*233*	*31*	*41*	*54*	*65*	*21*	*35*	*90*	*43*	*44*
7 or more	6.0	25.8	2.4	1.9	–	19.0	11.4	3.3	2.3	4.5
6 or more	7.7	35.5	4.9	1.9	–	19.0	14.3	6.7	2.3	4.5
5 or more	12.0	48.4	12.2	5.6	–	19.0	22.9	11.1	4.7	9.1
4 or more	18.9	61.3	26.8	7.4	1.5	33.3	34.3	17.8	11.6	9.1
3 or more	26.2	67.7	43.9	16.7	1.5	47.6	40.0	26.7	16.3	13.6
2 or more	38.6	87.1	51.2	35.2	6.2	66.7	57.1	40.0	20.9	25.0
1 or more	68.7	93.5	85.4	68.5	40.0	90.5	88.6	70.0	53.5	54.5
No low scores	31.3	6.5	14.6	31.5	60.0	9.5	11.4	30.0	46.5	45.5
≤5th percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	–	–	–	–	–	–	–	–	–	–
13 or more	–	–	–	–	–	–	–	–	–	–
12 or more	–	–	–	–	–	–	–	–	–	–
11 or more	–	–	–	–	–	–	–	–	–	–
10 or more	.9	–	–	–	–	4.8	5.7	–	–	–
9 or more	1.3	3.2	–	–	–	4.8	5.7	1.1	–	–
8 or more	1.7	6.5	–	–	–	4.8	5.7	2.2	–	–
7 or more	2.1	9.7	–	–	–	9.5	5.7	2.2	–	–
6 or more	4.3	19.4	–	–	–	19.0	5.7	2.2	2.3	2.3
5 or more	5.6	22.6	4.9	–	–	19.0	11.4	3.3	2.3	2.3
4 or more	9.0	32.3	7.3	3.7	–	19.0	25.7	5.6	2.3	4.5
3 or more	16.3	54.8	19.5	9.3	–	33.3	31.4	16.7	2.3	9.1
2 or more	26.6	71.0	39.0	16.7	3.1	38.1	48.6	27.8	14.0	13.6
1 or more	51.5	77.4	70.7	46.3	24.6	71.4	65.7	54.4	39.5	36.4
No low scores	48.5	22.6	29.3	53.7	75.4	28.6	34.3	45.6	60.5	63.6

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*233*	*31*	*41*	*54*	*65*	*21*	*35*	*90*	*43*	*44*
≤2nd percentile										
16 low scores	–	–	–	–	–	–	–	–	–	–
15 or more	–	–	–	–	–	–	–	–	–	–
14 or more	–	–	–	–	–	–	–	–	–	–
13 or more	–	–	–	–	–	–	–	–	–	–
12 or more	–	–	–	–	–	–	–	–	–	–
11 or more	–	–	–	–	–	–	–	–	–	–
10 or more	–	–	–	–	–	–	–	–	–	–
9 or more	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	.9	–	–	–	–	4.8	2.9	–	–	–
6 or more	.9	–	–	–	–	4.8	2.9	–	–	–
5 or more	3.4	16.1	–	–	–	14.3	5.7	2.2	–	2.3
4 or more	6.0	22.6	4.9	1.9	–	19.0	8.6	5.6	2.3	2.3
3 or more	9.4	35.5	12.2	1.9	–	19.0	20.0	8.9	2.3	4.5
2 or more	16.7	45.2	24.4	9.3	1.5	23.8	28.6	15.6	9.3	13.6
1 or more	39.1	74.2	46.3	35.2	15.4	57.1	51.4	41.1	32.6	22.7
No low scores	60.9	25.8	53.7	64.8	84.6	42.9	48.6	58.9	67.4	77.3

Note. All values represent cumulative percentages except for the rows labeled "No low scores," which provide the percentage of the normative sample with no scores falling under the low score cutoffs. See Table 19 for list of Total Achievement Scores for each D-KEFS Test. Abbreviations: 20Q = Twenty Questions Test; CWIT = Color-Word Interference Test; DF = Design Fluency Test; PT = Proverb Test; ST= Sorting Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency Test; Word Context Test = WC

Table 27. *Base Rates of Low Age-Adjusted D-KEFS Total Achievement Scores in 60-89 year-olds for the Four-Test Battery – 9 scores: TMT (1 score), VF (4 scores), CWIT (2 scores), TWT (2 scores)*

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*242*	*33*	*43*	*55*	*67*	*23*	*36*	*93*	*45*	*45*
≤25th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	2.5	6.1	–	–	–	4.3	5.6	3.2	–	–
6 or more	7.4	24.2	4.7	3.6	–	21.7	11.1	8.6	–	2.2
5 or more	15.7	36.4	23.3	5.5	3.0	26.1	22.2	16.1	6.7	13.3
4 or more	28.1	54.5	37.2	18.2	10.4	43.5	38.9	31.2	15.6	17.8
3 or more	43.4	81.8	55.8	30.9	22.4	60.9	66.7	46.2	22.2	31.1
2 or more	63.2	90.9	76.7	52.7	43.3	87.0	88.9	64.5	37.8	53.3
1 or more	86.8	100.0	90.7	87.3	74.6	100.0	100.0	86.0	73.3	84.4
No low scores	13.2	–	9.3	12.7	25.4	–	–	14.0	26.7	15.6
≤16th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	1.7	3.0	–	–	–	4.3	5.6	1.1	–	–
6 or more	1.7	3.0	–	–	–	4.3	5.6	1.1	–	–
5 or more	7.9	18.2	9.3	5.5	–	17.4	8.3	8.6	2.2	6.7
4 or more	15.7	30.3	20.9	12.7	3.0	21.7	25.0	18.3	8.9	6.7
3 or more	25.6	57.6	34.9	18.2	6.0	34.8	47.2	26.9	11.1	15.6
2 or more	43.0	81.8	48.8	34.5	19.4	73.9	61.1	43.0	24.4	31.1
1 or more	73.6	90.9	79.1	69.1	55.2	87.0	88.9	74.2	60.0	66.7
No low scores	26.4	9.1	20.9	30.9	44.8	13.0	11.1	25.8	40.0	33.3
≤9th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–

Number of Low Scores	Total Sample	WASI FSIQ ≤89	90-99	100-109	110+	Education ≤8	9-11	12	13-15	16+
N	*242*	*33*	*43*	*55*	*67*	*23*	*36*	*93*	*45*	*45*
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	.8	–	–	–	–	4.3	2.8	–	–	–
6 or more	.8	–	–	–	–	4.3	2.8	–	–	–
5 or more	3.7	9.1	2.3	3.6	–	8.7	5.6	2.2	2.2	4.4
4 or more	8.3	18.2	11.6	5.5	–	17.4	13.9	6.5	4.4	6.7
3 or more	15.3	36.4	20.9	12.7	3.0	26.1	19.4	14.0	11.1	13.3
2 or more	28.1	63.6	37.2	25.5	3.0	47.8	44.4	28.0	13.3	20.0
1 or more	57.4	84.8	60.5	56.4	34.3	82.6	77.8	53.8	44.4	48.9
No low scores	42.6	15.2	39.5	43.6	65.7	17.4	22.2	46.2	55.6	51.1
≤5th percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	.4	–	–	–	–	–	2.8	–	–	–
6 or more	.8	–	–	–	–	4.3	2.8	–	–	–
5 or more	1.2	3.0	–	–	–	4.3	2.8	1.1	–	–
4 or more	3.3	6.1	4.7	1.8	–	4.3	5.6	2.2	2.2	4.4
3 or more	7.0	18.2	11.6	1.8	–	17.4	13.9	5.4	2.2	4.4
2 or more	18.6	45.5	25.6	12.7	1.5	34.8	27.8	18.3	8.9	13.3
1 or more	43.8	72.7	51.2	36.4	20.9	69.6	61.1	43.0	28.9	33.3
No low scores	56.2	27.3	48.8	63.6	79.1	30.4	38.9	57.0	71.1	66.7
≤2nd percentile										
9 low scores	–	–	–	–	–	–	–	–	–	–
8 or more	–	–	–	–	–	–	–	–	–	–
7 or more	–	–	–	–	–	–	–	–	–	–
6 or more	.4	–	–	–	–	–	2.8	–	–	–
5 or more	.4	–	–	–	–	–	2.8	–	–	–

Number of Low Scores	Total Sample	WASI FSIQ				Education				
		≤89	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*242*	*33*	*43*	*55*	*67*	*23*	*36*	*93*	*45*	*45*
4 or more	.8	–	–	–	–	4.3	2.8	–	–	–
3 or more	4.1	9.1	7.0	–	–	8.7	8.3	4.3	2.2	–
2 or more	10.3	24.2	16.3	5.5	–	13.0	13.9	8.6	6.7	13.3
1 or more	32.6	63.6	44.2	25.5	10.4	52.2	47.2	32.3	20.0	24.4
No low scores	67.4	36.4	55.8	74.5	89.6	47.8	52.8	67.7	80.0	75.6

Note. All values represent cumulative percentages except for the rows labeled "No low scores," which provide the percentage of the normative sample with no scores falling under the low score cutoffs. See Table 19 for list of Total Achievement Scores for each D-KEFS Test. Abbreviations: CWIT = Color-Word Interference Test; TMT = Trail Making Test; TWT = Tower Test; VF = Verbal Fluency Test

Table 28. *Base Rates of Low Age-Adjusted D-KEFS Total Achievement Scores in 60-89 year-olds for the Three-Test Battery – 7 scores: TMT (1 score), VF (4 scores), CWIT (2 scores)*

		WASI FSIQ				Education				
Number of Low Scores	**Total Sample**	**≤89**	**90-99**	**100-109**	**110+**	**≤8**	**9-11**	**12**	**13-15**	**16+**
N	*337*	*57*	*61*	*78*	*89*	*37*	*53*	*127*	*65*	*55*
≤25th percentile										
7 low scores	.6	3.5	–	–	–	5.4	–	–	–	–
6 or more	3.6	10.5	1.6	1.3	–	13.5	3.8	3.1	–	1.8
5 or more	11.3	28.1	8.2	9.0	–	29.7	15.1	8.7	6.2	7.3
4 or more	20.8	52.6	23.0	12.8	3.4	37.8	34.0	18.9	9.2	14.5
3 or more	34.1	68.4	39.3	25.6	12.4	54.1	47.2	35.4	18.5	23.6
2 or more	51.3	87.7	59.0	38.5	29.2	75.7	69.8	52.0	30.8	40.0
1 or more	78.6	96.5	88.5	73.1	64.0	97.3	98.1	81.1	55.4	69.1
No low scores	21.4	3.5	11.5	26.9	36.0	2.7	1.9	18.9	44.6	30.9
≤16th percentile										
7 low scores	.6	3.5	–	–	–	5.4	–	–	–	–
6 or more	2.4	7.0	1.6	–	–	10.8	3.8	1.6	–	–
5 or more	4.5	14.0	1.6	3.8	–	18.9	5.7	2.4	–	3.6
4 or more	11.9	28.1	13.1	6.4	–	24.3	17.0	11.8	6.2	5.5
3 or more	21.4	50.9	21.3	14.1	3.4	37.8	34.0	19.7	12.3	12.7
2 or more	34.4	71.9	37.7	26.9	11.2	56.8	50.9	33.1	20.0	23.6
1 or more	64.1	89.5	70.5	59.0	40.4	86.5	79.2	65.4	44.6	54.5
No low scores	35.9	10.5	29.5	41.0	59.6	13.5	20.8	34.6	55.4	45.5
≤9th percentile										
7 low scores	–	–	–	–	–	–	–	–	–	–
6 or more	1.2	3.5	–	–	–	8.1	1.9	–	–	–
5 or more	2.4	5.3	1.6	2.6	–	10.8	1.9	.8	–	3.6
4 or more	6.8	19.3	8.2	3.8	–	21.6	7.5	5.5	3.1	3.6
3 or more	14.2	36.8	18.0	7.7	1.1	32.4	17.0	12.6	9.2	9.1

Number of Low Scores	Total Sample	WASI FSIQ ≤89	WASI FSIQ 90-99	WASI FSIQ 100-109	WASI FSIQ 110+	Education ≤8	Education 9-11	Education 12	Education 13-15	Education 16+
N	*337*	*57*	*61*	*78*	*89*	*37*	*53*	*127*	*65*	*55*
2 or more	26.1	59.6	27.9	23.1	3.4	45.9	37.7	25.2	12.3	20.0
1 or more	49.0	77.2	55.7	47.4	22.5	73.0	67.9	45.7	32.3	41.8
No low scores	51.0	22.8	44.3	52.6	77.5	27.0	32.1	54.3	67.7	58.2
≤5th percentile										
7 low scores	–	–	–	–	–	–	–	–	–	–
6 or more	.3	–	–	–	–	–	1.9	–	–	–
5 or more	.6	–	–	–	–	2.7	1.9	–	–	–
4 or more	3.0	7.0	3.3	2.6	–	8.1	3.8	3.1	–	1.8
3 or more	8.0	19.3	13.1	5.1	–	21.6	7.5	7.1	4.6	5.5
2 or more	17.8	45.6	18.0	11.5	2.2	35.1	24.5	15.0	9.2	16.4
1 or more	39.2	64.9	49.2	34.6	14.6	59.5	54.7	40.2	20.0	30.9
No low scores	60.8	35.1	50.8	65.4	85.4	40.5	45.3	59.8	80.0	69.1
≤2nd percentile										
7 low scores	–	–	–	–	–	–	–	–	–	–
6 or more	–	–	–	–	–	–	–	–	–	–
5 or more	.3	–	–	–	–	–	1.9	–	–	–
4 or more	.6	–	–	1.3	–	–	3.8	–	–	–
3 or more	3.6	8.8	6.6	1.3	–	13.5	3.8	3.9	–	–
2 or more	10.7	26.3	13.1	6.4	–	18.9	11.3	10.2	4.6	12.7
1 or more	29.1	50.9	37.7	24.4	7.9	45.9	37.7	29.9	15.4	23.6
No low scores	70.9	49.1	62.3	75.6	92.1	54.1	62.3	70.1	84.6	76.4

Note. All values represent cumulative percentages except for the rows labeled "No low scores," which provide the percentage of the normative sample with no scores falling under the low score cutoffs. See Table 19 for list of Total Achievement Scores for each D-KEFS Test. Abbreviations: CWIT = Color-Word Interference Test; TMT = Trail Making Test; VF = Verbal Fluency Test

Bibliography

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. doi:10.1016/j.bandc.2012.01.005

Agostino, A., Johnson, J., & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: Problem type matters. Journal *of Experimental Child Psychology, 105*(4), 286-305. doi:10.1016/j.jecp.2009.09.006

Aichert, D. S., Wöstmann, N. M., Costa, A., Macare, C., Wenig, J. R., Möller, H., & ... Ettinger, U. (2012). Associations between trait impulsivity and prepotent response inhibition. *Journal of Clinical and Experimental Neuropsychology, 34*(10), 1016-1032. doi:10.1080/13803395.2012.706261

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. doi:10.1109/TAC.1974.1100705

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., ... & Snyder, P. J. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 270-279. doi:10.1016/j.jalz.2011.03.008

Allan, N. P., & Lonigan, C. J. (2011). Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. *Developmental Psychology, 47*(4), 905-915. doi:10.1037/a0023748

Allan, N. P., & Lonigan, C. J. (2014). Exploring dimensionality of effortful control using hot and cool tasks in a sample of preschool children. *Journal of Experimental Child Psychology, 122*, 33-47. doi:10.1016/j.jecp.2013.11.013

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

Aran-Filippetti, V. (2013). Structure and invariance of executive functioning tasks across socioeconomic status: evidence from spanish-speaking children. *The Spanish Journal of Psychology, 16*, 1-15. doi:10.1017/sjp.2013.102

Aron, A. R. (2008). Progress in executive-function research: From tasks to functions to regions to networks. *Current Directions in Psychological Science, 17*(2), 124-129.

Baddeley, A., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.) *Recent advances in learning and motivation* (pp. 47–89). New York: Academic Press.

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

Barkley, R. A. (2001). The executive functions and self-regulation: An evolutionary neuropsychological perspective. *Neuropsychology review*, *11*(1), 1-29. doi:10.1023/A:1009085417776

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved*. New York: Guilford Press.

Baron, I. S. (2004). Delis-Kaplan Executive Function System. *Child Neuropsychology, 10*(2), 147-152. doi:10.1080/09297040490911140

Barroso, F. (1983). An approach to the study of attentional components in auditory tasks. *Journal of Auditory Research, 23*(3), 157-180.

Bell-McGinty, S., Podell, K., Franzen, M., Baird, A. D., & Williams, M. J. (2002). Standard measures of executive function in predicting instrumental activities of daily living in older adults. *International Journal of Geriatric Psychiatry, 17*(9), 828-834. doi:10.1002/gps.646

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606. doi:10.1037/0033-2909.88.3.588

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

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

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: “Abnormal” neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology, 24,* 31-46. doi:10.1093/arclin/acn001

Booth, J. N., Boyle, J. M., & Kelly, S. W. (2010). Do tasks make a difference? Accounting for heterogeneity of performance of children with reading difficulties on tasks of executive function: Findings from a meta-analysis. *British Journal of Developmental Psychology, 28*(1), 133-176. doi:10.1348/026151009X485432

Box, G. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. Oxford: John Wiley & Sons.

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. doi:10.1002/icd.1871

Brooks , B. L. (2010). Seeing the forest for the trees: Prevalence of low scores on the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV), *Psychological Assessment, 22*(3), 650-656. doi:10.1037/a0019781

Brooks, B. L., Holdnack, J. A., & Iverson, G. L. (2011). Advanced clinical interpretation of the WAIS-IV and WMS-IV: Prevalence of low scores varies by level of intelligence and years of education. *Assessment, 18*(2), 156-167. doi:10.1177/1073191110385316

Brooks, B. L., Iverson, G. L., & White, T. (2007). Substantial risk of "Accidental MCI" in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society, 13*(3), 490-500. doi:10.10170S1355617707070531

Brooks, B. L., Iverson, G. L., Feldman, H. H., & Holdnack, J. A. (2009). Minimizing misdiagnosis: Psychometric criteria for possible or probable memory impairment. *Dementia and Geriatric Cognitive Disorders, 27*(5), 439-450. doi:10.1159/000215390

Brooks, B. L., Iverson, G. L., Holdnack, J. A., & Feldman, H. H. (2008). Potential for misclassification of mild cognitive impairment: A study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International Neuropsychological Society, 14*(3), 463-478. doi:10.10170S1355617708080521

Brooks, B. L., Iverson, G. L., Koushik, N. S., Mazur-Mosiewicz, A., Horton Jr, A. M., & Reynolds, C. R. (2013b). Prevalence of low scores in children and adolescents on the test of verbal conceptualization and fluency. *Applied Neuropsychology: Child, 2*(1), 70-77. doi:10.1080/21622965.2012.742792

Brooks, B. L., Iverson, G. L., Lanting, S. C., Horton, A. M. Jr., & Reynolds, C. R. (2012). Improving test interpretation for detecting executive dysfunction in adults and older adults: Prevalence of low scores on the Test of Verbal Conceptualization and Fluency. *Applied Neuropsychology, 19*(1), 61-70. doi:10.1080/09084282.2012.651951

Brooks, B. L., Sherman, E. M., & Iverson, G. L. (2010). Healthy children get low scores too: prevalence of low scores on the NEPSY-II in preschoolers, children, and adolescents. *Archives of Clinical Neuropsychology, 25*(3), 182-190. doi:10.1093/arclin/acq005

Brooks, B.L, & Iverson, G.L. (2012). Improving the accuracy when identifying cognitive impairment in pediatric neuropsychological assessments. In E. Sherman and B. Brooks (eds.), *Pediatric Forensic Neuropsychology*, pp. 66-88. New York: Oxford University Press.

Brooks, B.L., Iverson, G.L., & Holdnack, J.A. (2013a). Understanding and using multivariate base rates with the WAIS-IV/WMS-IV. In Holdnack, J.A., Drozdick, L.W., Weiss, L.G., & Iverson, G.L. (Eds.), *WAIS-IV/WMS-IV/ACS: Advanced clinical interpretation*. San Diego, CA: Elsevier Science.

Brown, T. E. (2008). ADD/ADHD and impaired executive function in clinical practice. *Current Psychiatry Reports, 10*(5), 407-411. doi:10.1007/s11920-008-0065-7

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

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. doi:10.1016/j.intell.2012.05.006

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 79-113). Hove, UK: Psychology Press.

Cahn-Weiner, D. A., Boyle, P. A., & Malloy, P. F. (2002). Tests of executive function predict instrumental activities of daily living in community-dwelling older individuals. *Applied Neuropsychology*, *9*(3), 187-191. doi:10.1207/S15324826AN0903_8

Canivez, G. L. (2014). Construct validity of the WISC-IV with a referred sample: Direct versus indirect hierarchical structures. *School Psychology Quarterly*, *29*(1), 38-51. doi:10.1037/spq0000032

Canivez, G. L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247-271). Gottingen, Germany: Hogrefe Publishing.

Canivez, G. L., & Kush, J. C. (2013). WISC–IV and WAIS–IV structural validity: Alternate methods, alternate results. Commentary on Weiss et al. (2013a) and Weiss et al. (2013b). Journal of Psychoeducational Assessment, 31, 157–169. doi:10.1177/0734282913478036

Carlson, S. M., White, R. E., & Davis-Unger, A. C. (2014). Evidence for a relation between executive function and pretense representation in preschool children. *Cognitive Development, 29*, 1-16. doi:10.1016/j.cogdev.2013.09.001

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

Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology, 21*(3), 217-227. doi:10.1016/j.acn.2005.12.002

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189-225. doi:10.1207/s15327906mbr4102_5

Cheung, G.W., & Rensvold, R.B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. doi:10.1207/S15328007SEM0902_5

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

Chudley, A. E., Conry, J., Cook, J. L., Loock, C., Rosales, T., & LeBlanc, N. (2005). Fetal alcohol spectrum disorder: Canadian guidelines for diagnosis. *Canadian Medical Association Journal, 172*(5 Supplement), S1-S21. doi:10.1503/cmaj.1040302

Cinan, S., Özen, G., & Hampshire, A. (2013). Confirmatory factor analysis on separability of planning and insight constructs. *Journal of Cognitive Psychology, 25*(1), 7-23. doi:10.1080/20445911.2012.729035

Cirino, P. T., Chapieski, L. M., & Massman, P. J. (2000). Card sorting performance and ADHD symptomatology in children and adolescents with Tourette syndrome. *Journal of Clinical and Experimental Neuropsychology, 22*(2), 245-256. doi:10.1076/1380-3395(200004)22:2;1-1;FT245

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155-159. doi:10.1037/0033-2909.112.1.155

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. doi:10.1016/j.neuroscience.2005.05.035

Collette, F., van der Linden, M., Laureys, S., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2005). Exploring the unity and diversity of the neural substrates of executive functioning. *Human brain mapping*, *25*(4), 409-423. doi:10.1002/hbm.20118

Crawford, J. R., Garthwaite , P. H., & Gault , C. B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology, 21*(4), 419-430. doi:10.1037/0894-4105.21.4.419

Crawford, J. R., Garthwaite, P. H., Sutherland, D., & Borland, N. (2011). Some supplementary methods for the analysis of the Delis−Kaplan Executive Function System. *Psychological Assessment, 23*(4), 888-898. doi:10.1037/a0023712

de Frias, C. M., Dixon, R. A., & Strauss, E. (2006). Structure of four executive functioning tests in healthy older adults. *Neuropsychology*, *20*(2), 206-214. doi: 10.1037/0894-4105.20.2.206

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. doi:10.1037/a0016743

Deckel, A. W., & Hesselbrock, V. (1996). Behavioral and Cognitive Measurements Predict Scores on the MAST: A 3-Year Prospective Study. *Alcoholism: Clinical and Experimental Research*, *20*(7), 1173-1178. doi:10.1111/j.1530-0277.1996.tb01107.x

Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using shared variance techniques with normal or mixed clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society*, *9*, 936–946.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan Executive Function System*. Odessa, FL: Psychological Assessment Resources.

Delis, D. C., Kramer, J. H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan Executive Function System: an update. *Journal of the International Neuropsychological Society, 10*(2), 301-303. doi:10.1017/S1355617704102191

Della Sala, S., Gray, C., Spinnler, H., & Trivelli, C. (1998). Frontal lobe functioning in man: The riddle revisited. *Archives of Clinical Neuropsychology, 13*(8), 663-682. doi:10.1016/S0887-6177(97)00093-0

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959-964. doi:10.1126/science.1204529

Duan, X., Wei, S., Wang, G., & Shi, J. (2010). The relationship between executive functions and intelligence on 11-to 12-year-old children. *Psychological Test and Assessment Modeling*, *52*(4), 419-431.

Duggan, E. C., & Garcia-Barrera, M. A. (2015). Executive functioning and intelligence. In S. Goldstein, J. A. Naglieri, & D. Princiotta (Eds.), *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts* (pp. 435-458). New York: Springer Publishing Co. doi:10.1007/978-1-4939-1562-0_27

Edwards, P., Clarke, M., DiGuiseppi, C., Pratap, S., Roberts, I., & Wentz, R. (2002). Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Statistics in Medicine*, *21*(11), 1635-1640. doi:10.1002/sim.1190

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*(3), 430–457. doi:10.1207/S15328007SEM0803_5

Eslinger, P. J. (1996). Conceptualizing, describing, and measuring components of executive function. In G. R. Lyons & N. A. Krasnegor (Eds.), *Attention, memory and executive function* (pp. 367-395). Baltimore, MD: Brooks Publishing.

Espy, K. A., Sheffield, T. D., Wiebe, S. A., Clark, C. A., & Moehr, M. J. (2011). Executive control and dimensions of problem behaviors in preschool children. *Journal of Child Psychology and Psychiatry*, *52*(1), 33-46. doi:10.1111/j.1469-7610.2010.02265.x

Ettenhofer, M. L., Hambrick, D. Z., & Abeles, N. (2006). Reliability and stability of executive functioning in older adults. *Neuropsychology*, *20*(5), 607-613. doi: 10.1037/0894-4105.20.5.607

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*(1), 56-83. doi:10.1080/10705519909540119

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. doi:10.1111/jopy.12163

Floyd, R. G., Bergeron, R., Hamilton, G., & Parra, G. R. (2010). 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, 47*(7), 721-738. doi:10.1002/pits.20500

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. doi:10.1016/j.actpsy.2008.04.004

Frazier, D. T., Bettcher, B. M., Dutt, S., Patel, N., Mungas, D., Miller, J., ... & Kramer, J. H. (2015). Relationship between insulin-resistance processing speed and specific executive function profiles in neurologically intact older adults. *Journal of the International Neuropsychological Society*, *21*(8), 622-628. doi:10.1017/S1355617715000624

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

Friedman, N. P., Corley, R. P., Hewitt, J. K., & Wright, K. P. (2009). Individual differences in childhood sleep problems predict later cognitive executive control. *Sleep*, *32*(3), 323-333.

Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., & Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science, 18*(10), 893-900. doi:10.1111/j.1467-9280.2007.01997.x

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. doi:10.1037/dev0000075

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), 172-179. doi:10.1111/j.1467-9280.2006.01681.x

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. doi:10.1037/a0023750

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. doi:10.1037/0096-3445.137.2.201

Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*(1), 65-83. doi:10.1207/s15327906mbr4101_5

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31-60. doi:10.1037/0033-2909.134.1.31

Garza, J. P., Nelson, J. M., Sheffield, T. D., Choi, H., Clark, C. A., Wiebe, S. A., & Espy, K. A. (2014, February). Parsing executive control from foundational cognitive abilities in preschool: Application of the bifactor model to examine developmental change. In K. A. Espy (Chair), *The changing nature of executive control in preschool: Using statistical modeling to situate neuroscience in development.* Symposium conducted at the Forty-Second Annual Meeting of the International Neuropsychological Society, Seattle, Washington.

Gavett, B. E. (2015). The value of Bayes' theorem for interpreting abnormal test scores in cognitively healthy and clinical samples. *Journal of the International Neuropsychological Society*, *21*(3), 249-257. doi:10.1017/S1355617715000168

Greve, K. W., Stickle, T. R., Love, J. M., Bianchini, K. J., & Stanford, M. S. (2005). Latent structure of the Wisconsin Card Sorting Test: A confirmatory factor analytic study. *Archives of Clinical Neuropsychology*, *20*(3), 355-364. doi:10.1016/j.acn.2004.09.004

Grodzinsky, G. M., & Diamond, R. (1992). Frontal lobe functioning in boys with attention-deficit hyperactivity disorder. *Developmental Neuropsychology, 8*(4), 427-445. doi:10.1080/87565649209540536

Hancock, G. R. (2006). Power analysis in covariance structure modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 69-115). Greenwich, Connecticut: Information Age Publishing.

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

Homack, S., Lee, D., & Riccio, C. A. (2005). Test review: Delis-Kaplan executive function system. *Journal of Clinical and Experimental Neuropsychology, 27*(5), 599-609. doi:10.1080/13803390490918444

Hooper, D., Coughlan, J., Mullen, M. (2008) Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60

Hoyle, R.H., & Panter, A.T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158-176). Thousand Oaks, CA: Sage Publications.

Hoyme, H. E., May, P. A., Kalberg, W. O., Kodituwakku, P., Gossage, J. P., Trujillo, P. M., ... & Viljoen, D. L. (2005). A practical clinical approach to diagnosis of fetal alcohol spectrum disorders: Clarification of the 1996 institute of medicine criteria. *Pediatrics, 115*(1), 39-47. doi:10.1542/peds.2004-0259

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

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

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, *35*(1), 20-36. doi: 10.1080/87565640903325691

Huizinga, M., Dolan, C. V., & Van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017-2036. doi:10.1016/j.neuropsychologia.2006.01.010

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. doi:10.1037/0894-4105.22.4.508

Isquith, P. K., Roth, R. M., Gioia, G. A. (2009). *Tasks of Executive Control*. Lutz, Florida: Psychological Assessment Resources.

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. doi:10.1037/a0038557

Iverson, G.L. & Brooks, B.L. (2011). Improving accuracy for identifying cognitive impairment. In M.R. Schoenberg and J.G**.** Scott (Eds.), *The little black book of neuropsychology: A syndrome-based approach* (pp. 923-950). New York: Springer.

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

Kafer, K. L., & Hunter, M. (1997). On testing the face validity of planning/problem-solving tasks in a normal population. *Journal of the International Neuropsychological Society*, *3*(2), 108-119.

Kalechstein, A. D., van Gorp, W. G., & Rapport, L. J. (1998). Variability in clinical classification of raw test scores across normative data sets. *The Clinical Neuropsychologist, 12*(3), 339–347. doi:10.1076/clin.12.3.339.1991

Kamphaus, R. W. (2005). *Clinical assessment of child and adolescent intelligence* (2nd ed.). New York: Springer.

Karr, J. E., Areshenkoff, C. N., & Garcia-Barrera, M. A. (2014). The neuropsychological outcomes of concussion: A systematic review of meta-analyses on the cognitive sequelae of mild traumatic brain injury. *Neuropsychology, 28*(3), 321-336. doi:10.1037/neu0000037

Karr, J. E., Areshenkoff, C. N., Rast, P., & Garcia-Barrera, M. A. (2014). An empirical comparison of the therapeutic benefits of physical exercise and cognitive training on the executive functions of older adults: A meta-analysis of controlled trials. *Neuropsychology, 28*(6), 829-845. doi:10.1037/neu0000101

Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology Review, 18*(4), 320-338. doi:10.1007/s11065-008-9077-7

Keren-Happuch, E., Chen, S. A., Ho, M. R., & Desmond, J. E. (2014). A meta-analysis of cerebellar contributions to higher cognition from PET and fMRI studies. *Human Brain Mapping, 35*(2), 593-615. doi:10.1002/hbm.22194

Khoury, J. E., Milligan, K., & Girard, T. A. (2015). Executive functioning in children and adolescents prenatally exposed to alcohol: A meta-analytic review. *Neuropsychology Review, 25*(2), 149-170. doi:10.1007/s11065-015-9289-6

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

Kline, R. B. (2006). Reverse arrow dynamics: Formative measurement and feedback loops. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 43-68). Greenwich, Connecticut: Information Age Publishing.

Kramer, J. H., Mungas, D., Possin, K. L., Rankin, K. P., Boxer, A. L., Rosen, H. J., ... & Widmeyer, M. (2014). NIH EXAMINER: Conceptualization and development of an executive function battery. *Journal of the International Neuropsychological Society, 20*(1), 11-19. doi:10.1017/S1355617713001094

Krasny-Pacini, A., Chevignard, M., & Evans, J. (2014). Goal Management Training for rehabilitation of executive functions: A systematic review of effectiveness in patients with acquired brain injury. *Disability and Rehabilitation, 36*(2), 105-116, doi:10.3109/09638288.2013.777807

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. doi:10.1111/j.1467-9450.2011.00899.x

Latzman, R. D., & Markon, K. E. (2010). The factor structure and age-related factorial invariance of the Delis-Kaplan Executive Function System (D-KEFS). *Assessment*, *17*(2), 172-184. doi:10.1177/1073191109356254

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

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. doi:10.1111/j.2044-8279.2010.02016.x

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

Lerner, M. D., & Lonigan, C. J. (2014). Executive function among preschool children: Unitary versus distinct abilities. *Journal of Psychopathology and Behavioral Assessment, 36*(4), 626-639. doi:10.1007/s10862-014-9424-3

Levin, H. S., Fletcher, J. M., Kufera, J. A., Harward, H., Lilly, M. A., Mendelsohn, D., ... & Eisenberg, H. M. (1996). Dimensions of cognition measured by the tower of London and other cognitive tasks in head-injured children and adolescents. *Developmental Neuropsychology*, *12*(1), 17-34. doi:10.1080/87565649609540638

Lezak, M. D. (1976). *Neuropsychological assessment*. New York: Oxford University Press.

Lezak, M. D. (1982). The problem of assessing executive functions. *International Journal of Psychology*, *17*(1-4), 281-297. doi:10.1080/00207598208247445

Lezak, M. D. (1983). *Neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Lonie, J. A., Herrmann, L. L., Donaghey, C. L., & Ebmeier, K. P. (2008). Clinical referral patterns and cognitive profile in mild cognitive impairment. *The British Journal of Psychiatry*, *192*(1), 59-64. doi:10.1192/bjp.bp.107.035642

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904)"'General intelligence,' objectively determined and measured". *Journal of Personality and Social Psychology, 86*(1), 96-111. doi:10.1037/0022-3514.86.1.96

Luna, B., Marek, S., Larsen, B., Tervo-Clemmens, B., & Chahal, R. (2015). An integrative model of the maturation of cognitive control. *Annual Review of Neuroscience, 38*, 151-170. doi:10.1146/annurev-neuro-071714-034054

Luria, A. R. (1966). *Higher cortical functions in man*. New York: Basic Books.

MacDonald, S. W., Nyberg, L., & Bäckman, L. (2006). Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences, 29*(8), 474-480. doi:10.1016/j.tins.2006.06.011

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research how often do they really occur? *Perspectives on Psychological Science, 7*(6), 537-542. doi:10.1177/1745691612460688

Masten, A. S., Herbers, J. E., Desjardins, C. D., Cutuli, J. J., McCormick, C. M., Sapienza, J. K., ... & Zelazo, P. D. (2012). Executive function skills and school success in young children experiencing homelessness. *Educational Researcher, 41*(9), 375-384. doi:10.3102/0013189X12459883

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

McFall, G. P., Wiebe, S. A., Vergote, D., Jhamandas, J., Westaway, D., & Dixon, R. A. (2014). IDE (rs6583817) polymorphism and pulse pressure are independently and interactively associated with level and change in executive function in older adults. *Psychology And Aging, 29*(2), 418-430. doi:10.1037/a0034656

McFall, G. P., Wiebe, S. A., Vergote, D., Westaway, D., Jhamandas, J., & Dixon, R. A. (2013). IDE (rs6583817) polymorphism and type 2 diabetes differentially modify executive function in older adults. *Neurobiology Of Aging, 34*(9), 2208-2216. doi:10.1016/j.neurobiolaging.2013.03.010

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

Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development, 13*(3), 395-423. doi:10.1080/15248372.2011.585478

Miller, M. R., Müller, U., Giesbrecht, G. F., Carpendale, J. I., & Kerns, K. A. (2013). The contribution of executive function and social understanding to preschoolers' letter and math skills. *Cognitive Development, 28*(4), 331-349. doi:10.1016/j.cogdev.2012.10.005

Mistridis, P., Egli, S. C., Iverson, G. L., Berres, M., Willmes, K., Welsh-Bohmer, K. A., & Monsch, A. U. (2015). Considering the base rates of low performance in cognitively healthy older adults improves the accuracy to identify neurocognitive impairment with the consortium to establish a registry for Alzheimer's disease-neuropsychological assessment battery (CERAD-NAB). *European Archives of Psychiatry and Clinical Neuroscience, 265*(5), 407-417. doi:10.1007/s00406-014-0571-z

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. doi:10.1177/0963721411429458

Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language, 21*(2), 169-183. doi:10.1055/s-2000-7563

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

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*(7), 2693-2698. doi:10.1073/pnas.1010076108

Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*, 264–269. doi:10.7326/0003-4819-151-4-200908180-00135

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

Müller, U., & Kerns, K. (2015). The development of executive function. In L. S. Liben, U. Müller & R. M. Lerner (Eds.), *Handbook of child psychology and developmental science, Vol. 2: Cognitive processes* (7th ed., pp. 571-623). Hoboken, NJ: Wiley. doi:10.1002/9781118963418.childpsy214

Muthén, L, & Muthén, B. (2014). MPlus (Version 7.3) [Computer Software]. Los Angeles: Muthén & Muthén.

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. doi:10.3758/s13415-011-0083-5

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (pp. 1-18). New York: Plenum.

OCEBM (Oxford Centre for Evidence-Based Medicine) Levels of Evidence Working Group. (2011). *The Oxford Levels of Evidence 2*. Retrieved from http://www.cebm.net/index.aspx?o=5653.

Oosterlaan, J., Scheres, A., & Sergeant, J. A. (2005). Which executive functioning deficits are associated with AD/HD, ODD/CD and comorbid AD/HD+ ODD/CD? *Journal of Abnormal Child Psychology, 33*(1), 69-85. doi:10.1007/s10802-005-0935-y

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. Human *Brain Mapping, 25*(1), 46-59. doi:10.1002/hbm.20131

Packwood, S., Hodgetts, H. M., & Tremblay, S. (2011). A multiperspective approach to the conceptualization of executive functions. *Journal of Clinical and Experimental Neuropsychology*, *33*(4), 456-470. doi: 10.1080/13803395.2010.533157

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*(6), 531-536. doi:10.1177/1745691612463401

Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science a crisis of confidence? *Perspectives on Psychological Science, 7*(6), 528-530. doi:10.1177/1745691612465253

Pearson, Inc. (2001). *Standardization data from the Delis-Kaplan Executive Function System (D-KEFS)*. Copyright © 2001 NCS Pearson, Inc. Used with permission. All rights reserved. San Antonio: Pearson, Inc.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, *256*(3), 183-194. doi:10.1111/j.1365-2796.2004.01388.x

Petersen, R. C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., & Fratiglioni, L. (2014). Mild cognitive impairment: A concept in evolution. *Journal of Internal Medicine, 275*(3), 214-228. doi:10.1111/joim.12190

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

Phillips, L. H. (1997). Do ''frontal tests'' measure executive function? Issues of assessment and evidence from fluency tests. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 191–213). Hove, UK: Psychology Press.

Pickens, S., Ostwald, S. K., Murphy-Pace, K., & Bergstrom, N. (2010). Systematic review of current executive function measures in adults with and without cognitive impairments. *International Journal of Evidence-Based Healthcare, 8*(3), 110-125. doi:10.1111/j.1744-1609.2010.00170.x

Pribram, K. H. (1973). The primate frontal cortex – executive of the brain. In K. H. Pribram & A. R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 293-314). New York: Academic Press.

Psychological Corporation (1999). *Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: Psychological Corporation.

R Core Team. (2013). R: A language and environment for statistical computing [computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20,* 33-65. doi: 10.1016/j.acn.2004.02.005.

Rabin, L.A., Paolillo, E., & Barr, W.B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology*, 31, 206-230. doi:10.1093/arclin/acw007

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111-163. doi:10.2307/271063

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667-696.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544-559. doi:10.1080/00223891.2010.496477

Reynolds, C. R., & Horton, A. M. (2008). Assessing executive functions: A life-span perspective. *Psychology in the Schools*, *45*(9), 875-892. doi:10.1002/pits.20332

Roberts, N., & Grover, V. (2009). Theory development in information systems research using structural equation modeling: Evaluation and recommendations. In Y. K. Dwivedi (Ed.), Handbook of research on contemporary theoretical models in information systems (pp. 77-94). Hershey, Pennsylvania: Information Science Reference.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2011). Modeling a cascade of effects: The role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental Science*, *14*(5), 1161-1175. doi:10.1111/j.1467-7687.2011.01068.x

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012). Implications of infant cognition for executive functions at age 11. *Psychological Science, 23*(11), 1345-1355. doi:10.1177/0956797612444902

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1-36. doi:10.18637/jss.v048.i02

Royall, D. R., Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., ... & Coffey, C. E. (2002). Executive control function: a review of its promise and challenges for clinical research. A report from the Committee on Research of the American Neuropsychiatric Association. *The Journal of Neuropsychiatry and Clinical Neurosciences, 14*(4), 377-405. doi:10.1176/jnp.14.4.377

Sanchez-Cubillo, I., Perianez, J. A., Adrover-Roig, D., Rodriguez-Sanchez, J. M., Rios-Lago, M., Tirapu, J., & Barcelo, F. (2009). Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society, 15*(3), 438-450. doi:10.1017/S1355617709090626

Sanderson, S., Tatt, I. D., & Higgins, J. P. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International Journal of Epidemiology*, *36*(3), 666-676. doi:10.1093/ije/dym018

Sapkota, S., Vergote, D., Westaway, D., Jhamandas, J., & Dixon, R. A. (2015). Synergistic associations of catechol-O-methyltransferase and brain-derived neurotrophic factor with executive function in aging are selective and modified by apolipoprotein E. *Neurobiology of Aging, 36*(1), 249-256. doi:10.1016/j.neurobiolaging.2014.06.020

Scherer, R., & Tiemann, R. (2014). Evidence on the effects of task interactivity and grade level on thinking skills involved in complex problem solving. *Thinking Skills and Creativity*, *11*, 48-64. doi:10.1016/j.tsc.2013.10.003

Schmidt, M. (2003). Hit or miss? Insight into executive functions. *Journal of the International Neuropsychological Society*, *9*(6), 962–964. doi:10.1017/S1355617703230162

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323-338. doi:10.3200/JOER.99.6.323-338

Schretlen, D. J., Testa, S. M., Winicki, J. M., Pearlson, G. D., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society, 14*(3), 436-445. doi:10.1017/S1355617708080387.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464. doi:10.1214/aos/1176344136

Scott, J. C., Matt, G. E., Wrocklage, K. M., Crnich, C., Jordan, J., Southwick, S. M., ... & Schweinsburg, B. C. (2015). A quantitative meta-analysis of neurocognitive functioning in posttraumatic stress disorder. *Psychological Bulletin*, *141*(1), 105-140. doi:10.1037/a0038039

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.00772

Shunk, A. W., Davis, A. S., & Dean, R. S. (2006). Review of Delis-Kaplan Executive Function System (D-KEFS). *Applied Neuropsychology, 13*(4), 275-279. doi:10.1207/s15324826an1304_9

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366. doi:10.1177/0956797611417632

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin*, *139*(1), 81-132. doi:10.1037/a0028727

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*(5), 893-898. doi:10.1016/j.paid.2006.09.017

Sternberg, R. J., Grigorenko, E., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly*, *47*(1), 1-41. doi:10.1353/mpq.2001.0005

Stuss, D. T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology*, *53*(1), 401-433. doi:10.1146/annurev.psych.53.100901.135220

Swanson, J. (2005). The Delis-Kaplan Executive Function System: A Review. *Canadian Journal of School Psychology*, *20*(1/2), 117-128. doi:10.1177/0829573506295469

Tanaka, J. S. (1987). 'How big is big enough?': Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, *58*(1), 134-146. doi:10.1111/1467-8624.ep7264172

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

Teuber, H. L. (1972). Unity and diversity of frontal lobe functions. *Acta Neurobiologiae Experimentalis*, *32*, 615–656.

Thibeau, S., McFall, G. P., Wiebe, S. A., Anstey, K. J., & Dixon, R. A. (2016). Genetic factors moderate everyday physical activity effects on executive functions in aging: Evidence from the Victoria Longitudinal Study. *Neuropsychology*, *30*(1), 6-17. doi:10.1037/neu0000217

Tuleya, L. G. (Ed.). (2009). *Thesaurus of psychological index terms*. (Rev. 11th ed.). Washington, DC: American Psychological Association.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *The Quarterly Journal of Experimental Psychology*, *64*(3), 447-466. doi:10.1080/17470218.2010.505292

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. doi:10.1080/17405629.2013.840578

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

van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. (2013). The structure of executive functions in children: A closer examination of inhibition, shifting, and updating. *British Journal of Developmental Psychology, 31*(1), 70-87. doi:10.1111/j.2044-835X.2012.02079.x

Vandenbroucke, J. P., Von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., ... & Egger, M. (2007). Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. *Annals of internal medicine, 147*(8), 163-194. doi:10.7326/0003-4819-147-8-200710160-00010-w1

Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and Aging*, *25*(2), 343-355. doi: 10.1037/a0017729

Viterbori, P., Usai, M. C., Traverso, L., & De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: A longitudinal study. *Journal of Experimental Child Psychology*, *140*, 38-55. doi:10.1016/j.jecp.2015.06.014

Was, C. A. (2007). Further evidence that not all executive functions are equal. *Advances in Cognitive Psychology*, *3*(3), 399-407. doi: 10.2478/v10053-008-0004-5

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children—Fourth Edition among a national sample of referred students. *Psychological Assessment*, *22*(4), 782-787. doi:10.1037/a0020043

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2008). *Wechsler adult intelligence scale* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2014). *Wechsler intelligence scale for children* (5th ed.). San Antonio, TX: Pearson.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013a). WAIS-IV and clinical validation of the four-and five-factor interpretative approaches. *Journal of Psychoeducational Assessment*, *31*(2), 94-113. doi:10.1177/0734282913478030

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013b). WISC-IV and clinical validation of the four-and five-factor interpretative approaches. *Journal of Psychoeducational Assessment*, *31*(2), 114-131. doi:10.1177/0734282913478032

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental Neuropsychology, 4*(3), 199-230. doi:10.1080/87565648809540405

White, T., & Stern, R. A. (2003). *Neuropsychological assessment battery: Psychometric and technical manual*. Lutz, Florida: Psychological Assessment Resources.

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. doi:10.1037/0012-1649.44.2.575

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology, 108*(3), 436-452. doi:10.1016/j.jecp.2010.08.008

Willoughby, M. T., Blair, C. B., & The Family Life Project Investigators (2016). Measuring executive function in early childhood: A case for formative measurement. *Psychological Assessment*, *28*(3), 319-330. doi:10.1037/pas0000152

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M., & The Family Life Project Investigators (2010). The measurement of executive function at age 3 years: psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment 22*(2), 306-317. doi:10.1037/a0018708

Willoughby, M. T., Blair, C. B., Wirth, R. J., Greenberg, M., & The Family Life Project Investigators (2012a). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment*, *24*(1), 226. doi:10.1037/a0025361

Willoughby, M. T., Wirth, R. J., Blair, C. B., & The Family Life Project Investigators (2012b). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment, 24*(2), 418-431. doi:10.1037/a0025779

Willoughby, M., Holochwost, S. J., Blanton, Z. E., & Blair, C. B. (2014). Executive functions: Formative versus reflective measurement. *Measurement: Interdisciplinary Research & Perspectives, 12*(3), 69-95. doi:10.1080/15366367.2014.929453

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L. O., ... & Arai, H. (2004). Mild cognitive impairment–beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine, 256*(3), 240-246. doi:10.1111/j.1365-2796.2004.01380.x

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), e77770. doi:10.1371/journal.pone.0077770

Appendix A: Articles Excluded from Systematic Review Organized by Reason for Exclusion

1. Did not involve a sample or sub-sample of cognitively healthy participants

Barbey, A. K., Colom, R., & Grafman, J. (2013). Dorsolateral prefrontal contributions to human intelligence. *Neuropsychologia*, *51*(7), 1361-1369. doi:10.1016/j.neuropsychologia.2012.05.017

Brookshire, B., Levin, H. S., Song, J. X., & Zhang, L. (2004). Components of executive function in typically developing and head-injured children. *Developmental Neuropsychology*, *25*(1&2), 61-83. doi:10.1080/87565641.2004.9651922

Christopher, M. E., Miyake, A., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2012). Predicting word reading and comprehension with executive function and speed measures across development: a latent variable analysis. *Journal of Experimental Psychology: General*, *141*(3), 470-488. doi:10.1037/a0027375m

Harvey, D. J., Beckett, L. A., & Mungas, D. M. (2003). Multivariate modeling of two associated cognitive outcomes in a longitudinal study. *Journal of Alzheimer's Disease*, *5*(5), 357-365.

Schoemaker, K., Bunte, T., Wiebe, S. A., Espy, K. A., Deković, M., & Matthys, W. (2012). Executive function deficits in preschool children with ADHD and DBD. *Journal of Child Psychology and Psychiatry*, *53*(2), 111-119. doi:10.1111/j.1469-7610.2011.02468.x

2. Did not report a confirmatory factor analysis of a multidimensional measurement model of executive function

Aichert, D. S., Wöstmann, N. M., Costa, A., Macare, C., Wenig, J. R., Möller, H. J., Rubia, K., & Ettinger, U. (2012). Associations between trait impulsivity and prepotent response inhibition. *Journal of Clinical and Experimental Neuropsychology*, *34*(10), 1016-1032. doi:10.1080/13803395.2012.706261

Allan, N. P., & Lonigan, C. J. (2011). Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. *Developmental Psychology*, *47*(4), 905-915. doi:10.1037/a0023748

Allan, N. P., & Lonigan, C. J. (2014). Exploring dimensionality of effortful control using hot and cool tasks in a sample of preschool children. *Journal of Experimental Child Psychology*, *122*, 33-47. doi:10.1016/j.jecp.2013.11.013

Baena, E., Allen, P. A., Kaut, K. P., & Hall, R. J. (2010). On age differences in prefrontal function: the importance of emotional/cognitive integration. *Neuropsychologia*, *48*(1), 319-333. doi:10.1016/j.neuropsychologia.2009.09.021

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of ‘hot’and ‘cool’executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly*, *24*(3), 337-349. doi:10.1016/j.ecresq.2009.06.001

Brocki, K. C., & Bohlin, G. (2004). Executive functions in children aged 6 to 13: A dimensional and developmental study. *Developmental Neuropsychology*, *26*(2), 571-593. doi: 10.1207/s15326942dn2602_3

Chow, M., & Conway, A. R. (2015). The scope and control of attention: Sources of variance in working memory capacity. *Memory & Cognition*, *43*(3), 325-339. doi:10.3758/s13421-014-0496-9

Cinan, S., Özen, G., & Hampshire, A. (2013). Confirmatory factor analysis on separability of planning and insight constructs. *Journal of Cognitive Psychology*, *25*(1), 7-23. doi:10.1080/20445911.2012.729035

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

Decker, S. L., Hill, S. K., & Dean, R. S. (2007). Evidence of construct similarity in executive functions and fluid reasoning abilities. *International Journal of Neuroscience*, *117*(6), 735-748. doi:10.1080/00207450600910085

Egeland, J., Landrø, N. I., Tjemsland, E., & Walbækken, K. (2006). Norwegian norms and factor-structure of phonemic and semantic word list generation. *The Clinical Neuropsychologist*, *20*(4), 716-728. doi:10.1080/13854040500351008

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

Espy, K. A., Kaufmann, P. M., Glisky, M. L., & McDiarmid, M. D. (2001). New procedures to assess executive functions in preschool children. *The Clinical Neuropsychologist*, *15*(1), 46-58. doi:10.1076/clin.15.1.46.1908

Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology*, *26*(1), 465-486. doi:10.1207/s15326942dn2601_6

Ettenhofer, M. L., Hambrick, D. Z., & Abeles, N. (2006). Reliability and stability of executive functioning in older adults. *Neuropsychology*, *20*(5), 607-613. doi:10.1037/0894-4105.20.5.607

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology*, *26*(7), 874-890. doi:10.1080/13803390490510680

Floyd, R. G., Bergeron, R., Hamilton, G., & Parra, G. R. (2010). 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*, *47*(7), 721-738. doi:10.1002/pits.20500

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

Gansler, D. A., Suvak, M., Arean, P., & Alexopoulos, G. S. (2015). Role of executive dysfunction and dysexecutive behavior in late-life depression and disability. *The American Journal of Geriatric Psychiatry*, *23*(10), 1038-1045. doi:10.1016/j.jagp.2015.05.003

Gay, P., Rochat, L., Billieux, J., d'Acremont, M., & Van der Linden, M. (2008). Heterogeneous inhibition processes involved in different facets of self-reported impulsivity: Evidence from a community sample. *Acta Psychologica*, *129*(3), 332-339. doi:10.1016/j.actpsy.2008.08.010

Goschke, T. (2014). Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: advances, gaps, and needs in current research. *International Journal of Methods in Psychiatric Research*, *23*(S1), 41-57. doi:10.1002/mpr.1410

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement*, *36*(3), 189-213. doi:10.1177/0146621612439620

Gustafsson, J. E., & Wolff, U. (2015). Measuring fluid intelligence at age four. *Intelligence*, *50*, 175-185. doi:10.1016/j.intell.2015.04.008

Haring, L., Mõttus, R., Koch, K., Trei, M., & Maron, E. (2015). Factorial validity, measurement equivalence and cognitive performance of the Cambridge Neuropsychological Test Automated Battery (CANTAB) between patients with first-episode psychosis and healthy volunteers. *Psychological Medicine*, *45*(09), 1919-1929. doi:10.1017/S0033291714003018

Howard, S. J., Johnson, J., & Pascual-Leone, J. (2014). Clarifying inhibitory control: Diversity and development of attentional inhibition. *Cognitive Development*, *31*, 1-21. doi:10.1016/j.cogdev.2014.03.001

Howard, S. J., & Okely, A. D. (2015). Catching fish and avoiding sharks: Investigating factors that influence developmentally appropriate measurement of preschoolers' inhibitory control. *Journal of Psychoeducational Assessment*, *33*(6), 585-596. doi:10.1177/0734282914562933

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, *35*(1), 20-36. doi:10.1080/87565640903325691

Karasinski, C. (2015). Language ability, executive functioning and behaviour in school-age children. *International Journal of Language & Communication Disorders*, *50*(2), 144-150. doi:10.1111/1460-6984.12104

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3-to 12-year-old Finnish children. *Developmental Neuropsychology*, *20*(1), 407-428. doi:10.1207/S15326942DN2001_6

Latzman, R. D., & Markon, K. E. (2010). The factor structure and age-related factorial invariance of the Delis-Kaplan Executive Function System (D-KEFS). *Assessment*, *17*(2), 172-184. doi:10.1177/1073191109356254

Lee, C. S., & Therriault, D. J. (2013). The cognitive underpinnings of creative thought: A latent variable analysis exploring the roles of intelligence and working memory in three creative thinking processes. *Intelligence*, *41*(5), 306-320. doi:10.1016/j.intell.2013.04.008

Lin, F., Roiland, R., Chen, D. G. D., & Qiu, C. (2015). Linking cognition and frailty in middle and old age: Metabolic syndrome matters. *International Journal of Geriatric Psychiatry*, *30*(1), 64-71. doi:10.1002/gps.4115

Mann, S. L., Selby, E. A., Bates, M. E., & Contrada, R. J. (2015). Integrating affective and cognitive correlates of heart rate variability: A structural equation modeling approach. *International Journal of Psychophysiology*, *98*(1), 76-86. doi:10.1016/j.ijpsycho.2015.07.003

Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive Development*, *33*, 40-55. doi:10.1016/j.cogdev.2014.07.001

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, *24*(12), 2351-2360. doi:10.1177/0956797613495418

McAuley, T., & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of Experimental Child Psychology*, *108*(3), 453-468. doi:10.1016/j.jecp.2010.08.009

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

Mitchell, M. B., Shaughnessy, L. W., Shirk, S. D., Yang, F. M., & Atri, A. (2012). Neuropsychological test performance and cognitive reserve in healthy aging and the Alzheimer's disease spectrum: A theoretically driven factor analysis. *Journal of the International Neuropsychological Society*, *18*(06), 1071-1080. doi:10.1017/S1355617712000859

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621-640. doi:10.1037//0096-3445.130.4.621

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. doi:10.1177/0963721411429458

Nieuwenhuis, S., & Monsell, S. (2002). Residual costs in task switching: Testing the failure-to-engage hypothesis. *Psychonomic Bulletin & Review*, *9*(1), 86-92. doi:10.3758/BF03196259

Pineda, D. A., & Merchan, V. (2003). Executive function in young Colombian adults. *International Journal of Neuroscience*, *113*(3), 397-410. doi:10.1080/00207450390162164

Rodríguez-Aranda, C., & Sundet, K. (2006). The frontal hypothesis of cognitive aging: Factor structure and age effects on four frontal tests among healthy individuals. *The Journal of Genetic Psychology*, *167*(3), 269-287. doi:10.3200/GNTP.167.3.269-287

Roebers, C. M., Röthlisberger, M., Neuenschwander, R., Cimeli, P., Michel, E., & Jäger, K. (2014). The relation between cognitive and motor performance and their relevance for children’s transition to school: a latent variable approach. *Human Movement Science*, *33*, 284-297. doi:10.1016/j.humov.2013.08.011

Royall, D. R., & Palmer, R. F. (2014). "Executive functions" cannot be distinguished from general intelligence: Two variations on a single theme within a symphony of latent variance. *Frontiers in Behavioral Neuroscience*, *8*, 369, 1-10. doi:10.3389/fnbeh.2014.00369

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*(4), 566-594. doi:10.1037/0096-3445.132.4.566

Scherer, R., & Tiemann, R. (2014). Evidence on the effects of task interactivity and grade level on thinking skills involved in complex problem solving. *Thinking Skills and Creativity*, *11*, 48-64. doi:10.1016/j.tsc.2013.10.003

Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental Neuropsychology*, *26*(1), 445-464. doi:10.1207/s15326942dn2601_5

Shahabi, S. R., Abad, F. J., & Colom, R. (2014). Short-term storage is a stable predictor of fluid intelligence whereas working memory capacity and executive function are not: A comprehensive study with Iranian schoolchildren. *Intelligence*, *44*, 134-141. doi:10.1016/j.intell.2014.04.004

Spiess, M. A., Meier, B., & Roebers, C. M. (2015). Prospective memory, executive functions, and metacognition are already differentiated in young elementary school children: Evidence from latent factor modeling. *Swiss Journal of Psychology, 74*(4), 229-241. doi:10.1024/1421-0185/a000165

St. Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, *59*(4), 745-759. doi:10.1080/17470210500162854

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

Tucker-Drob, E. M., & Salthouse, T. A. (2009). Methods and measures: Confirmatory factor analysis and multidimensional scaling for construct validation of cognitive abilities. *International Journal of Behavioral Development*, *33*(3), 277-285. doi:10.1177/0165025409104489

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica*, *134*(1), 16-28. doi:10.1016/j.actpsy.2009.11.010

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2011). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *The Quarterly Journal of Experimental Psychology*, *64*(3), 447-466. doi:10.1080/17470218.2010.505292

Visu-Petra, L., Benga, O., & Miclea, M. (2007). Dimensions of attention and executive functioning in 5-to 12-years-old children: Neuropsychological assessment with the NEPSY battery. *Cognition, Brain, Behavior*, *11*(3), 585-608.

Visu-Petra, L., Cheie, L., Benga, O., & Miclea, M. (2012). The structure of executive functions in preschoolers: An investigation using the NEPSY battery. *Procedia-Social and Behavioral Sciences, 33*, 627-631. doi:10.1016/j.sbspro.2012.01.197

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, *102*(1), 43-53. doi:10.1037/a0016738

Yehene, E., & Meiran, N. (2007). Is there a general task switching ability? *Acta Psychologica*, *126*(3), 169-195. doi:10.1016/j.actpsy.2006.11.007

Ziegler, G., Dahnke, R., Winkler, A. D., & Gaser, C. (2013). Partial least squares correlation of multivariate cognitive abilities and local brain structure in children and adolescents. *NeuroImage*, *82*, 284-294. doi:10.1016/j.neuroimage.2013.05.088

3. Included fewer than two indicators, deriving from separate tests, per construct evaluated

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence, 46*, 73-83. doi:10.1016/j.intell.2014.05.007

Cassidy, A. R. (2016). Executive function and psychosocial adjustment in healthy children and adolescents: A latent variable modelling investigation. *Child Neuropsychology*, *22*(3), 292-317. doi:10.1080/09297049.2014.994484

Shing, Y. L., Lindenberger, U., Diamond, A., Li, S. C., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. Developmental Neuropsychology, 35(6), 679-697. doi:10.1080/87565641.2010.508546

Wu, K. K., Chan, S. K., Leung, P. W. L., Liu, W., Leung, F. L. T., & Ng, R. (2011). Components and developmental differences of executive functioning for school-aged children. *Developmental Neuropsychology*, *36*(3), 319-337. doi:10.1080/87565641.2010.549979

4. Included indicators that were not derived from performance-based cognitive or neuropsychological tests

Brydges, C. R., Fox, A. M., Reid, C. L., & Anderson, M. (2014). Predictive validity of the N2 and P3 ERP components to executive functioning in children: A latent-variable analysis. *Frontiers in Human Neuroscience*, *8*(80), 1-10. doi:10.3389/fnhum.2014.00080

Fuhs, M. W., & Day, J. D. (2011). Verbal ability and executive functioning development in preschoolers at head start. *Developmental Psychology*, *47*(2), 404-416. doi:10.1037/a0021065

Samyn, V., Roeyers, H., Bijttebier, P., Rosseel, Y., & Wiersema, J. R. (2015). Assessing effortful control in typical and atypical development: Are questionnaires and neuropsychological measures interchangeable? A latent-variable analysis. *Research in Developmental Disabilities*, *36*, 587-599. doi:10.1016/j.ridd.2014.10.018

5. Included indicators from fewer than three separate cognitive or neuropsychological test in their measurement model

Ardila, A., & Pineda, D. A. (2000). Factor structure of nonverbal cognition. *International Journal of Neuroscience*, *104*(1), 125-144. doi:10.3109/00207450009035013

Engel de Abreu, P. M., & Gathercole, S. E. (2012). Executive and phonological processes in second-language acquisition. *Journal of Educational Psychology*, *104*(4), 974-986. doi:10.1037/a0028390

Thomas, M. L., Brown, G. G., Gur, R. C., Moore, T. M., Patt, V. M., Nock, M. K., Naifeh, J. A., Heeringa, S., Ursano, R. J., Stein, M. B., & on behalf of the Army STARRS Collaborators. (2015). Measurement of latent cognitive abilities involved in concept identification learning. *Journal of Clinical and Experimental Neuropsychology*, *37*(6), 653-669. doi:10.1080/13803395.2015.1042358

Yeh, Z. (2013). Role of theory of mind and executive function in explaining social intelligence: A structural equation modeling approach. *Aging & Mental Health*, *17*(5), 527-534. doi:10.1080/13607863.201 .72 58235

6. Shared Dataset with other article included in the systematic review

Brydges, C. R., Fox, A. M., Reid, C. L., & Anderson, M. (2014). The differentiation of executive functions in middle and late childhood: a longitudinal latent-variable analysis. *Intelligence*, *47*, 34-43. doi:10.1016/j.intell.2014.08.010

Bull, R., Espy, K. A., Wiebe, S. A., Sheffield, T. D., & Nelson, J. M. (2011). Using confirmatory factor analysis to understand executive control in preschool children: Sources of variation in emergent mathematic achievement. *Developmental Science*, *14*(4), 679-692. doi:10.1111/j.1467-7687.2010.01012.x

de Frias, C. M., Dixon, R. A., & Strauss, E. (2006). Structure of four executive functioning tests in healthy older adults. *Neuropsychology*, *20*(2), 206. doi:10.1037/0894-4105.20.2.206

Espy, K. A., Sheffield, T. D., Wiebe, S. A., Clark, C. A., & Moehr, M. J. (2011). Executive control and dimensions of problem behaviors in preschool children. *Journal of Child Psychology and Psychiatry*, *52*(1), 33-46. doi:10.1111/j.1469-7610.2010.02265.x

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172-179. doi:10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., & Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science*, *18*(10), 893-900. doi:10.1111/j.1467-9280.2007.01997.x

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. doi:10.1037/dev0000075

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. doi:10.1037/0096-3445.137.2.201

Friedman, N. P., Corley, R. P., Hewitt, J. K., & Wright, K. P. (2009). Individual differences in childhood sleep problems predict later cognitive executive control. *Sleep*, *32*(3), 323-333.

Hughes, C., & Ensor, R. (2011). Individual differences in growth in executive function across the transition to school predict externalizing and internalizing behaviors and self-perceived academic success at 6 years of age. *Journal of Experimental Child Psychology*, *108*(3), 663-676. doi:10.1016/j.jecp.2010.06.005

McFall, G. P., Wiebe, S. A., Vergote, D., Westaway, D., Jhamandas, J., & Dixon, R. A. (2013). IDE (rs6583817) polymorphism and type 2 diabetes differentially modify executive function in older adults. *Neurobiology of Aging*, *34*(9), 2208-2216. doi:10.1016/j.neurobiolaging.2013.03.010

McFall, G. P., Wiebe, S. A., Vergote, D., Jhamandas, J., Westaway, D., & Dixon, R. A. (2014). IDE (rs6583817) polymorphism and pulse pressure are independently and interactively associated with level and change in executive function in older adults. *Psychology and Aging*, *29*(2), 418-430. doi:10.1037/a0034656

Miller, M. R., Müller, U., Giesbrecht, G. F., Carpendale, J. I., & Kerns, K. A. (2013). The contribution of executive function and social understanding to preschoolers' letter and math skills. *Cognitive Development*, *28*(4), 331-349. doi:10.1016/j.cogdev.2012.10.005

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2011). Modeling a cascade of effects: The role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental Science*, *14*(5), 1161-1175. doi:10.1111/j.1467-7687.2011.01068.x

Sapkota, S., Vergote, D., Westaway, D., Jhamandas, J., & Dixon, R. A. (2015). Synergistic associations of catechol-O-methyltransferase and brain-derived neurotrophic factor with executive function in aging are selective and modified by apolipoprotein E. *Neurobiology of Aging*, *36*(1), 249-256. doi:10.1016/j.neurobiolaging.2014.06.020

Thibeau, S., McFall, G. P., Wiebe, S. A., Anstey, K. J., & Dixon, R. A. (2016). Genetic factors moderate everyday physical activity effects on executive functions in aging: Evidence from the Victoria Longitudinal Study. *Neuropsychology*, *30*(1), 6-17. doi:10.1037/neu0000217

Van der Ven, S. H., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2012). The development of executive functions and early mathematics: A dynamic relationship. *British Journal of Educational Psychology*, *82*(1), 100-119. doi:10.1111/j.2044-8279.2011.02035.x

Viterbori, P., Usai, M. C., Traverso, L., & De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: A longitudinal study. *Journal of Experimental Child Psychology*, *140*, 38-55. doi:10.1016/j.jecp.2015.06.014

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2010). The measurement of executive function at age 3 years: psychometric properties and criterion validity of a new battery of tasks. *Psychological Assessment*, *22*(2), 306-317. doi: 10.1037/a0018708

Willoughby, M., & Blair, C. (2011). Test-retest reliability of a new executive function battery for use in early childhood. *Child Neuropsychology*, *17*(6), 564-579. doi:10.1080/09297049.2011.554390

Willoughby, M. T., Wirth, R. J., & Blair, C. B. (2012). Executive function in early childhood: Longitudinal measurement invariance and developmental change. *Psychological Assessment*, *24*(2), 418-431. doi:10.1037/a0025779

Willoughby, M. T., Blair, C. B., & The Family Life Project Investigators (2015). Measuring executive function in early childhood: A case for formative measurement. *Psychological Assessment*, *28*(3), 319-330. doi:10.1037/pas0000152

Appendix B: Syntax for Re-Analysis of Executive Function Test Battery Correlation Matrices

Directory Structure

The following directory structure should be used so the correct syntax is included in the correct folders for the model to run properly.

- R_Code_Simulation
    - Correlation_Matrices
        - Agostino 2010
            - Agostino2010.txt
        - AranFilippetti 2013
            - AranFilippetti2013.txt
        - Etc.
    - Papers_Syntax
        - paper_Agostino2010.R
        - paper_AranFilippetti2013.R
        - Etc.
    - Sim_code
        - bootstrapCor.R
        - importCorMatrix.R
        - modelFit.R
        - runAndSave.R
    - Simulation_data
    - Run_simulation.R

Correlation Matrices

Each of the following pages shows the means (top row), standard deviations (second row), and inter-test correlations (all subsequent rows) for the test batteries of each study included in the re-analysis of confirmatory factor analyses. If the means and standard deviations are missing, they were replaced with 999. The re-analysis disregarded the means and standard deviations and analyzed solely the correlation matrices because not all studies reported means and standard deviations.

Agostino2010.txt

3.75	4.42	4.39	0.71	13.79	8.47	47.4	0.86	26.37	0.36
0.72	1.03	1.37	0.18	6.35	4.81	16.26	0.11	20.81	0.17
1									
0.45	1								
0.36	0.37	1							
0.22	0.26	0.32	1						
0.22	0.12	0.24	0.28	1					
0.34	0.1	0.2	0.35	0.35	1				
0.51	0.35	0.38	0.37	0.35	0.38	1			
0.22	0.17	0.2	0.24	0.12	0.2	0.32	1		
0.4	0.23	0.31	0.24	0.21	0.19	0.56	0.23	1	
0.32	0.27	0.47	0.32	0.29	0.2	0.47	0.19	0.36	1

AranFilippetti2013.txt

999 999 999 999 999 999 999 999 999
999 999 999 999 999 999 999 999 999

1								
0.69	1							
0.73	0.62	1						
0.44	0.28	0.37	1					
0.58	0.47	0.51	0.62	1				
0.65	0.57	0.57	0.27	0.41	1			
0.64	0.53	0.56	0.35	0.49	0.57	1		
0.65	0.54	0.56	0.29	0.41	0.59	0.66	1	
0.4	0.32	0.38	0.28	0.35	0.32	0.31	0.33	1

Brydges2012.txt

999 999 999 999 999 999 999 999 999 999 999 999 999
999 999 999 999 999 999 999 999 999 999 999 999 999

1												
0.03	1											
0.13	0.06	1										
0.32	-.03	0.17	1									
0.26	0.13	0.07	0.36	1								
0.14	0.03	0.06	0.36	0.22	1							
0.23	0.05	0.13	0.38	0.2	0.17	1						
0.39	0.02	0.18	0.4	0.31	0.29	0.22	1					
0.29	-.01	0.2	0.4	0.28	0.13	0.28	0.26	1				
0.29	0.03	0.19	0.48	0.31	0.24	0.41	0.38	0.43	1			
0.37	-.04	0.18	0.45	0.32	0.32	0.39	0.41	0.42	0.59	1		
0.28	-.10	0.12	0.45	0.25	0.63	0.31	0.4	0.3	0.4	0.52	1	
0.28	-.10	0.14	0.56	0.35	0.52	0.32	0.48	0.36	0.45	0.52	0.79	1

deFrias2009CE.txt

5.9	1.46	5.35	109.67	3.63	4.09
1.13	0.48	1.79	19.48	0.99	1.51
1					
0.07	1				
-0.15	0.09	1			
0.19	0.12	0.07	1		
0.06	0.22	0.04	0.19	1	
0.06	0.14	0.11	0.29	0.52	1

deFrias2009CN.txt

5.8	1.18	4.93	94.16	2.9	3.12
1.26	0.62	2.18	27	0.94	1.13
1					
0.04	1				
0.18	0.05	1			
0.06	0.07	0.24	1		
0.19	0.14	0.18	0.21	1	
0.07	0.12	0.15	0.18	0.49	1

Duan2010.txt

```
980.544	995.69	7	20.75	568.45	593.05	49.33
374.76	280.52	7.2	17.77	338.9	517.62	7.24
1
0.748	1
0.202	-0.146	1
0.243	0.032	0.359	1
0.647	0.525	0.216	0.343	1
0.377	0.304	-0.064	0.141	0.481	1
-0.548	-0.393	-0.343	-0.302	-0.337	-0.203	1
```

Fleming2016.txt

0.71	250	132	0.79	0.9	0.96	259	217	146
0.2	38 69	0.15	0.24	0.14	180	180	115	
1								
0.09	1							
0.17	0.11	1						
0.05	0.08	0.02	1					
0.29	0.06	0.15	0.38	1				
0.24	0	0.08	0.2	0.31	1			
0.18	0.05	0.08	0.02	0.03	0.07	1		
0.18	-.04	0.08	0.05	0.05	-.01	0.37	1	
0.18	0.15	0.08	0.08	0.08	0.06	0.39	0.4	1

Friedman2011.txt

```
1.04	282	214	0.94	1.09	1.17	331	331	333	102
0.2	63	90	0.18	0.25	0.17	183	189	181	11
1
0.26	1
0.16	0.14	1
0.18	0.23	0.21	1
0.25	0.16	0.25	0.46	1
0.21	0.24	0.12	0.27	0.26	1
0.14	0.26	0.23	0.13	0.19	0.15	1
0.16	0.22	0.27	0.14	0.14	0.12	0.41	1
0.21	0.28	0.25	0.17	0.16	0.19	0.48	0.43	1
0.23	0.22	0.26	0.54	0.43	0.3	0.05	0.09	0.17	1
```

Hedden2006.txt

0.62	8.11	41.98	3.46	8.31	12.27	0.73	0.12	10.76	13.23	14.99	5.11	6.35	0.57	4.07	26.99	11.05	8.12	0.79	0.27	0.67	0.65
0.4	3.96	18.74	2.14	2.58	1.66	0.16	0.1	3.91	3.93	4.06	1.64	1.59	0.24	1.72	5.09	2.83	2.04	0.16	0.27	0.22	0.24
1																					
0.23	1																				
0.16	0.28	1																			
0.09	0.16	0.23	1																		
0.22	0.28	0.32	0.35	1																	
0.19	0.28	0.26	0.27	0.18	1																
0.2	0.13	0.2	0.17	0.14	0.22	1															
0.06	-0.07	0.21	-0.06	0.07	0.15	0.15	1														
0.15	0.21	0.21	0.34	0.23	0.27	0.14	0.05	1													
0.1	0.2	0.16	0.22	0.08	0.02	0.14	0.03	0.42	1												
0.17	0.01	0.07	0.28	0.23	0.19	0.17	0.11	0.35	0.35	1											
0.23	0.31	0.22	0.31	0.16	0.25	0.22	0.06	0.26	0.26	0.29	1										
0.23	0.21	0.21	0.22	0.1	0.21	0.12	0.04	0.15	0.15	0.24	0.73	1									
0.18	0.27	0.08	0.21	0.09	0.22	0.02	-0.1	0.11	0.12	0.31	0.47	0.49	1								
0.22	0.25	0.11	0.22	0.09	0.28	0.09	-0.07	0.08	0.1	0.26	0.48	0.42	0.76	1							
0.28	0.09	0.25	0.27	0.12	0.31	0.41	0.21	0.2	0.11	0.28	0.32	0.32	0.19	0.17	1						
0.12	0.04	0.16	0.14	0.1	0.25	0.25	0.2	0.14	0.08	0.15	0.16	0.32	0.13	0.05	0.46	1					
0.24	0.12	0.15	0.04	0.16	0.18	0.08	0.03	0.09	0	0.02	0.19	0.22	0.18	0.16	0.24	0.23	1				
0.18	0.22	0.2	0.26	0.26	0.27	0.04	-0.12	0.22	0.23	0.31	0.36	0.33	0.43	0.35	0.15	0.14	0.14	1			
-0.28	-0.13	-0.06	-0.16	-0.25	-0.1	-0.14	0.01	-0.16	-0.38	-0.23	-0.25	-0.16	-0.08	-0.12	-0.01	-0.04	-0.09	-0.46	1		
0.27	0.12	0.22	0.18	0.32	0.24	0.2	0	0.29	0.27	0.38	0.35	0.36	0.14	0.19	0.11	0.08	0.19	0.65	-0.93	1	
-0.1	-0.15	0.06	-0.03	-0.02	0.06	0.02	0.22	-0.1	-0.28	-0.05	0	0.08	-0.04	-0.06	0.12	0.16	-0.03	-0.24	0.56	-0.29	1

Hull2008.txt

```
202.22	217.12	41.02	6.55	11.28	13.7	24.11	243.68	63.22	0.28	87.67	6.56	0.04	52.28	21.7	60.24
167.2	166.05	10.07	5.67	8.41	9.37	12.82	102.69	60.91	0.15	24.48	4.66	0.02	7.87	2.85	5.58
1
0.49	1
0.07	-0.1	1
-0.22	-0.1	0.45	1
-0.14	0.06	0.30	0.42	1
-0.18	0.05	0.06	0.4	0.45	1
-0.05	-0.05	0.17	0.4	0.21	0.4	1
-0.14	-0.11	0.24	0.24	0.16	0.2	-0.03	1
-0.01	0.02	0.05	-0.06	0.19	0.16	-0.12	0.27	1
0.16	0.26	0.02	0.18	0.15	0.26	-0.22	0.13	0.09	1
-0.34	-0.05	-0.03	0.43	0.25	0.45	0.22	0.2	0.02	-0.01	1
-0.19	0.06	-0.01	0.36	0.36	0.39	0.19	0.3	0	0.24	0.43	1
-0.29	-0.25	0.09	0.32	0.58	0.33	-0.01	0.14	0.03	0.53	0.01	0.47	1
-0.04	-0.15	-0.27	-0.51	-0.42	-0.56	-0.35	-0.13	0.03	-0.22	-0.2	-0.37	-0.32	1
-0.07	-0.05	-0.23	-0.62	-0.52	-0.29	-0.15	-0.31	-0.11	-0.22	-0.34	-0.29	-0.32	0.36	1
-0.02	0.08	0.22	0.01	0.05	-0.02	-0.24	0.1	0.14	0.53	-0.18	0.04	0.37	-0.28	0	1
```

Ito2015.txt

61.98	247	141	73.32	75.12	80.73	243	241	156
14.36	32	70	8.6	13.49	6.8	156	160	119
1								
0.16	1							
0.18	0.09	1						
0.19	0.03	0.17	1					
0.32	0.02	0.3	0.37	1				
0.39	0.04	0.14	0.31	0.38	1			
0.18	0.11	0.23	0.07	0.12	0.11	1		
0.07	0	0.12	0.02	0.06	0.06	0.37	1	
0.25	0.1	0.16	0.08	0.12	0.11	0.5	0.41	1

Klauer2010Study1.txt

626	226	121	1.15	0.87	3.66	4.16
439	166	92	0.16	0.23	0.75	1.63
1						
0.18	1					
0.05	0.31	1				
0.12	0.12	0.13	1			
0.1	-0.04	0.17	0.26	1		
0.03	0.08	0.05	0.21	0.27	1	
-0.02	-0.02	0.05	0.11	-0.06	0.24	1

Klauer2010Study2.txt

```
172	232	134	83	135	58	29	9.87	9.54	11.75
108	123	94	40	1163	28	22	1.48	1.41	1.62
1
0.15	1
0.3	0.09	1
0	0.17	0.21	1
0.07	0.05	0.15	0.23	1
-0.16	0	-0.05	0.05	0.3	1
-0.01	0.12	0.12	0.25	0.17	0.21	1
-0.09	-0.12	-0.02	0.26	-0.1	0.04	0.13	1
-0.16	0.01	-0.06	0.22	-0.05	0.13	0.17	0.65	1
-0.04	0.01	0.04	0.22	-0.02	0.13	0.18	0.59	0.51	1
```

Lehto2003.txt

```
23 48.7 27.4 20.8 3.8 46.4 43.4 53.3 25.1 6 35.3 35.7 8 14.3
8.7 21.2 11.9 11.6 3.7 9.6 9.9 14 3.4 1.3 15.7 3.6 1.8 7.8
1
0.41 1
0.64 0.52 1
0.14 -0.11 0.13 1
-0.09 0.1 0.04 -0.51 1
-0.08 -0.05 0.07 0.06 -0.02 1
-0.05 -0.14 -0.02 0.13 -0.15 0.65 1
-0.12 -0.34 -0.27 0.22 -0.31 0.14 0.26 1
0.04 -0.2 -0.06 0.33 -0.36 0.25 0.36 0.25 1
-0.08 -0.22 -0.18 0.06 -0.06 0.18 0.2 0.25 0.31 1
0.08 0.18 0.13 -0.34 0.34 -0.24 -0.28 -0.27 -0.43 -0.36 1
0.05 0.17 0.13 -0.16 0.11 -0.13 -0.05 -0.23 -0.13 -0.16 0.61 1
-0.15 -0.26 -0.13 0.38 -0.37 0.08 0.2 0.27 0.29 0.16 -0.28 -0.11 1
0.18 0.36 0.22 -0.4 0.33 -0.09 -0.22 -0.39 -0.31 -0.24 0.31 0.12 -0.85 1
```

Miller2012.txt

0.81	1.14	0.73	7.8	6.24	0.79	15.86	0.07	2.39	0.75	0.74	0.79	12.15	67.87
1.14	1.25	0.13	6.39	7.67	0.18	3.46	0.04	2.03	0.22	0.2	0.2	4.15	18.24
1													
0.78	1												
0.23	0.24	1											
0.38	0.5	0.19	1										
0.2	0.26	0.09	0.28	1									
0.29	0.35	0.13	0.58	0.9	1								
0.12	0.26	0.19	0.15	0.24	0.24	1							
0.18	0.18	0.15	0.2	0.3	0.35	0.09	1						
0.13	0.15	0.15	0.24	0.39	0.42	0.31	0.17	1					
0.23	0.31	0.15	0.37	0.23	0.33	0.2	0.19	0.43	1				
0.3	0.41	0.12	0.48	0.34	0.47	0.28	0.17	0.18	0.22	1			
0.18	0.22	0.1	0.29	0.17	0.27	0.16	0.28	0.14	0.27	0.25	1		
0.45	0.59	0.19	0.44	0.09	0.22	0.19	0.13	0.19	0.4	0.4	0.26	1	
0.46	0.48	0.18	0.29	0.35	0.39	0.22	0.29	0.3	0.36	0.33	0.26	0.53	1

Miyake2000.txt

```
15.5 546 210 0.63 0.7 0.99 1.16 0.78 166 32 46 0 0 43 0.89
10.8 250 160 0.14 0.26 0.13 0.16 0.29 60 12 12 1 1 6 0.13
1
0.32 1
0.23 0.32 1
0.23 0.08 0.12 1
0.22 0.19 0 0.15 1
0.24 0.11 0.21 0.34 0.27 1
0.15 0.17 0.11 0.12 0.26 0.22 1
0.11 0.13 0.06 0.1 0.09 0.04 0.19 1
0.07 0.09 -0.05 0.11 0.16 0.18 0.2 0.18 1
0.26 0.13 0.18 0.09 0.19 0.14 0.15 -0.01 0.1 1
0.08 0.1 -0.09 0.13 0.18 0.14 0.21 0.08 0.17 -0.02 1
0.2 0.13 0.01 0.03 0.11 0.19 0.24 0.12 0.11 0.13 0.1 1
0.2 -0.07 0.07 0.29 0.06 0.19 0.02 0.18 0.01 -0.08 0.12 0.02 1
0.09 0.08 -0.04 0.41 0.28 0.34 0.16 0.13 0.2 0.16 0.04 0.17 0.13 1
-0.03 -0.02 0.05 -0.09 -0.03 0.12 -0.08 -0.16 0.06 0.06 -0.18 -0.05 -0.09 -0.14 1
```

Rose2012.txt

999	999	999	999	999	999	999	999	999	999
999	999	999	999	999	999	999	999	999	999
1									
0.3	1								
0.19	0.31	1							
0.39	0.13	0.21	1						
0.22	0.1	0.21	0.24	1					
0.4	0.23	0.23	0.37	0.39	1				
0.17	0.01	0.35	0.17	0.1	0.16	1			
0.19	0.22	0.36	0.13	0.2	0.1	0.26	1		
0.23	0.19	0.31	0.22	0.28	0.25	0.04	0.3	1	
0.22	0.06	0.18	0.11	0.17	0.1	0.01	0.17	0.36	1

Usai2014.txt

0.55	22.98	4.63	2.04	27.91	0.86	17.68
0.23	4.35	2.71	0.59	6.49	0.25	3.39
1						
0.252	1					
0.182	0.158	1				
0.146	0.327	0.245	1			
-0.02	0.046	0.154	0.156	1		
0.156	0.134	0.07	0.371	0.112	1	
0.023	0.136	0.27	0.173	-0.029	0.095	1

Xu2013Ages7to9.txt

86.21	58.64	69.99	50.21	415.47	782.88	611.46
10.58	13.68	12.36	15.62	190.62	307.07	328.82
1						
0.36	1					
0.19	0.27	1				
0.16	0.25	0.28	1			
0.12	0.06	0.15	0.13	1		
0.1	0.2	0.26	0.24	0.14	1	
0.21	0.28	0.32	0.21	0.12	0.18	1

Xu2013Ages10to12.txt

91.12	66.3	80.9	58.42	309.48	722.03	591.55
6.36	13.83	10.21	18.06	185.82	284.28	284.37
1						
0.3	1					
0.15	0.36	1				
0.17	0.21	0.14	1			
0.19	0.24	0.2	0.14	1		
0.25	0.29	0.23	0.19	0.19	1	
0.14	0.11	0.15	0.23	0.04	0.11	1

Xu2013Ages13to15.txt

93.41	71.22	86.66	69.37	230.29	608.42	543.05
4.69	13.77	9.81	17.85	115.01	236.76	289.69
1						
0.32	1					
0.24	0.3	1				
0.34	0.27	0.16	1			
0.2	0.16	0.2	0.25	1		
0.02	0.18	0.15	0.28	0.33	1	
0.18	0.22	0.17	0.25	0.1	0.21	1

Syntax for Re-Analysis

Each of the following pages lists the syntax used to specify the competing measurement models evaluated in the re-analysis. The variables are all named x1, x2, x3, etc. and were not given their specific variable name in these scripts. The number for each of these labels (i.e., the 1 in x1) corresponds to the row of the correlation matrix for this variable in its respective study.

paper_Agostino2010.R

```
# Location
folder = 'Agostino 2010'
file = 'Agostino2010'

# Paper data
N = 155
vars = c(4:10) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General        =~ x4 + x5 + x6 + x7 + x8 + x9 + x10
'

multimodel <- '
inhibit        =~ x4 + x5 + x6
update         =~ x7 + x8
shift          =~ x9 + x10
'

shiupdmerge <- '
inhibit        =~ x4 + x5 + x6
shiupd         =~ x7 + x8 + x9 + x10
'

inhupdmerge <- '
shift          =~ x9 + x10
inhupd         =~ x4 + x5 + x6 + x7 + x8
'

inhshimerge <- '
update         =~ x7 + x8
inhshi         =~ x4 + x5 + x6 + x9 + x10
'

bifactor <- '
general        =~ x4 + x5 + x6 + x7 + x8 + x9 + x10
inhibit        =~ x4 + x5 + x6
update         =~ v1*x7 + v1*x8
```

```
shift          =~ v2*x9 + v2*x10
general        ~~ 0*shift
general        ~~ 0*update
general        ~~ 0*inhibit
shift         ~~ 0*update
shift         ~~ 0*inhibit
update         ~~ 0*inhibit
'

binoinh <- '
general         =~ x4 + x5 + x6 + x7 + x8 + x9 + x10
update          =~ v1*x7 + v1*x8
shift          =~ v2*x9 + v2*x10
general        ~~ 0*shift
general        ~~ 0*update
shift         ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = bin
```

paper_AranFilippetti2013.R

```
# Location
folder = 'AranFilippetti 2013'
file = 'AranFilippetti2013'

# Paper data
N = 248
vars = c(1:9) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
'

multimodel <- '
update       =~ x1 + x2 + x3
shift        =~ x4 + x5 + x6
inhibit      =~ x7 + x8 + x9
'

shiupdmerge <- '
shiupd       =~ x1 + x2 + x3 + x4 + x5 + x6
inhibit      =~ x7 + x8 + x9
'

inhupdmerge <- '
shift        =~ x4 + x5 + x6
inhupd       =~ x1 + x2 + x3 + x7 + x8 + x9
'

inhshimerge <- '
update       =~ x1 + x2 + x3
inhshi       =~ x4 + x5 + x6 + x7 + x8 + x9
'

bifactor <- '
general      =~ x1 + x2 + x3  + x4 + x5 + x6 + x7 + x8 + x9
update       =~ x1 + x2 + x3
shift        =~ x4 + x5 + x6
```

```
inhibit        =~ x7 + x8 + x9
general        ~~ 0*shift
general        ~~ 0*update
general        ~~ 0*inhibit
shift          ~~ 0*update
shift          ~~ 0*inhibit
update         ~~ 0*inhibit
'

binoinh <- '
general        =~ x1 + x2 + x3  + x4 + x5 + x6 + x7 + x8 + x9
update         =~ x1 + x2 + x3
shift          =~ x4 + x5 + x6
general        ~~ 0*shift
general        ~~ 0*update
shift          ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Brydges2012.R

```
# Location
folder = 'Brydges 2012'
file = 'Brydges2012'

# Paper data
N = 215
vars = c(1:9) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models

uniModel <- '
General       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
'

multimodel <- '
inhibit     =~ x1 + x2 + x3
update      =~ x4 + x5 + x6
shift       =~ x7 + x8 + x9
'

shiupdmerge <- '
inhibit     =~ x1 + x2 + x3
shiupd      =~ x4 + x5 + x6 + x7 + x8 + x9
'

inhupdmerge <- '
inhupd      =~ x1 + x2 + x3 + x4 + x5 + x6
shift       =~ x7 + x8 + x9
'

inhshimerge <- '
inhshi      =~ x1 + x2 + x3 + x7 + x8 + x9
update      =~ x4 + x5 + x6
'

bifactor <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
inhibit       =~ x1 + x2 + x3
update        =~ x4 + x5 + x6
```

```
shift        =~ x7 + x8 + x9
general       ~~ 0*shift
general       ~~ 0*update
general       ~~ 0*inhibit
shift        ~~ 0*update
shift        ~~ 0*inhibit
update        ~~ 0*inhibit
'

binoinh <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
update        =~ x4 + x5 + x6
shift        =~ x7 + x8 + x9
general       ~~ 0*shift
general       ~~ 0*update
shift        ~~ 0*update
'


# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_deFrias2009CE.R

```
# Location
folder = 'de Frias 2009'
file = 'deFrias2009CE'

# Paper data
N = 77
vars = c(1:6) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General       =~ x1 + x2 + x3 + x4 + x5 + x6
'

multimodel <- '
inhibit       =~ x1 + x2
shift         =~ x3 + x4
update        =~ x5 + x6
'

shiupdmerge <- '
inhibit      =~ x1 + x2
shiupd       =~ x3 + x4 + x5 + x6
'

inhupdmerge <- '
inhupd       =~ x1 + x2  + x5 + x6
shift        =~ x3 + x4
'

inhshimerge <- '
inhshi       =~ x1 + x2 + x3 + x4
update       =~ x5 + x6
'

bifactor <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6
inhibit       =~ v1*x1 + v1*x2
shift         =~ v2*x3 + v2*x4
update        =~ v3*x5 + v3*x6
```

```
general       ~~ 0*inhibit
general       ~~ 0*shift
general       ~~ 0*update
inhibit       ~~ 0*shift
inhibit       ~~ 0*update
shift         ~~ 0*update
'

binoinh <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6
shift         =~ v4*x3 + v4*x4
update        =~ v5*x5 + v5*x6
general       ~~ 0*shift
general       ~~ 0*update
shift         ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_deFrias2009CN.R

```
# Location
folder = 'de Frias 2009'
file = 'deFrias2009CN'

# Paper data
N = 276
vars = c(1:6) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General        =~ x1 + x2 + x3 + x4 + x5 + x6
'

multimodel <- '
inhibit        =~ x1 + x2
shift          =~ x3 + x4
update         =~ x5 + x6
'

shiupdmerge <- '
inhibit      =~ x1 + x2
shiupd       =~ x3 + x4 + x5 + x6
'

inhupdmerge <- '
inhupd       =~ x1 + x2  + x5 + x6
shift        =~ x3 + x4
'

inhshimerge <- '
inhshi       =~ x1 + x2 + x3 + x4
update       =~ x5 + x6
'

bifactor <- '
general        =~ x1 + x2 + x3 + x4 + x5 + x6
inhibit        =~ v1*x1 + v1*x2
```

```
shift       =~ v2*x3 + v2*x4
update      =~ v3*x5 + v3*x6
general     ~~ 0*inhibit
general     ~~ 0*shift
general     ~~ 0*update
inhibit     ~~ 0*shift
inhibit     ~~ 0*update
shift       ~~ 0*update
'

binoinh <- '
general     =~ x1 + x2 + x3 + x4 + x5 + x6
shift       =~ v4*x3 + v4*x4
update      =~ v5*x5 + v5*x6
general     ~~ 0*shift
general     ~~ 0*update
shift       ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Duan2010.R

```
# Location
folder = 'Duan 2010'
file = 'Duan2010'

# Paper data
N = 61
vars = c(1:6) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General        =~ x1 + x2 + x3 + x4 + x5 + x6
'

multimodel <- '
update         =~ x1 + x2
inhibit        =~ x3 + x4
shift          =~ x5 + x6
'

shiupdmerge <- '
inhibit      =~ x3 + x4
shiupd       =~ x1 + x2 + x5 + x6
'

inhupdmerge <- '
inhupd       =~ x1 + x2  + x3 + x4
shift        =~ x5 + x6
'

inhshimerge <- '
update       =~ x1 + x2
inhshi       =~ x3 + x4 + x5 + x6
'

bifactor <- '
general        =~ x1 + x2 + x3 + x4 + x5 + x6
update         =~ v1*x1 + v1*x2
inhibit        =~ v2*x3 + v2*x4
shift          =~ v3*x5 + v3*x6
```

```
general       ~~ 0*inhibit
general       ~~ 0*shift
general       ~~ 0*update
inhibit       ~~ 0*shift
inhibit       ~~ 0*update
shift         ~~ 0*update
'

binoinh <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6
update        =~ v1*x1 + v1*x2
shift         =~ v3*x5 + v3*x6
general       ~~ 0*shift
general       ~~ 0*update
shift         ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Fleming2016.R

```
# Location
folder = 'Fleming 2016'
file = 'Fleming2016'

# Paper data
N = 420
vars = c(1:9) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                        vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models

unimodel <- '
General       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
'

multimodel <- '
inhibit     =~ x1 + x2 + x3
update      =~ x4 + x5 + x6
shift       =~ x7 + x8 + x9
'

shiupdmerge <- '
inhibit     =~ x1 + x2 + x3
shiupd      =~ x4 + x5 + x6 + x7 + x8 + x9
'

inhupdmerge <- '
inhupd      =~ x1 + x2 + x3 + x4 + x5 + x6
shift       =~ x7 + x8 + x9
'

inhshimerge <- '
inhshi      =~ x1 + x2 + x3 + x7 + x8 + x9
update      =~ x4 + x5 + x6
'

bifactor <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
inhibit       =~ x1 + x2 + x3
update        =~ x4 + x5 + x6
```

```
shift         =~ x7 + x8 + x9
general         ~~ 0*shift
general         ~~ 0*update
general         ~~ 0*inhibit
shift         ~~ 0*update
shift         ~~ 0*inhibit
update          ~~ 0*inhibit
'

binoinh <- '
general         =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
update          =~ x4 + x5 + x6
shift         =~ x7 + x8 + x9
general         ~~ 0*shift
general         ~~ 0*update
shift         ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Friedman2011.R

```
# Location
folder = 'Friedman 2011'
file = 'Friedman2011'

# Paper data
N = 813
vars = c(1:9) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                        vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)


# Specify Models
unimodel <- '
General        =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
'

multimodel <- '
shift          =~ x1 + x2 + x3
update         =~ x4 + x5 + x6
inhibit        =~ x7 + x8 + x9
'

shiupdmerge <- '
shiupd         =~ x1 + x2 + x3 + x4 + x5 + x6
inhibit        =~ x7 + x8 + x9
'

inhupdmerge <- '
shift          =~ x1 + x2 + x3
inhupd         =~ x4 + x5 + x6 + x7 + x8 + x9
'

inhshimerge <- '
update         =~ x4 + x5 + x6
inhshi         =~ x7 + x8 + x9 + x1 + x2 + x3
'

bifactor <- '
general        =~ x1 + x2 + x3  + x4 + x5 + x6 + x7 + x8 + x9
shift          =~ x1 + x2 + x3
update         =~ x4 + x5 + x6
```

```
inhibit      =~ x7 + x8 + x9
general      ~~ 0*shift
general      ~~ 0*update
general      ~~ 0*inhibit
shift        ~~ 0*update
shift        ~~ 0*inhibit
update       ~~ 0*inhibit
'

binoinh <- '
general      =~ x1 + x2 + x3  + x4 + x5 + x6 + x7 + x8 + x9
shift        =~ x1 + x2 + x3
update       =~ x4 + x5 + x6
general      ~~ 0*shift
general      ~~ 0*update
shift        ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Hedden2006.R

```
# Location
folder = 'Hedden 2006'
file = 'Hedden2006'

# Paper data
N = 121
vars = c(1:8) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                        vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)


# Specify Models
unimodel <- '
General        =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8
'

multimodel <- '
shift          =~ x1 + x2 + x3
update         =~ x4 + x5 + x6
inhibit        =~ x7 + x8
'

shiupdmerge <- '
shiupd         =~ x1 + x2 + x3 + x4 + x5 + x6
inhibit        =~ x7 + x8
'

inhupdmerge <- '
shift          =~ x1 + x2 + x3
inhupd         =~ x4 + x5 + x6 + x7 + x8
'

inhshimerge <- '
update         =~ x4 + x5 + x6
inhshi         =~ x7 + x8 + x1 + x2 + x3
'

bifactor <- '
general        =~ x1 + x2 + x3  + x4 + x5 + x6 + x7 + x8
shift          =~ x1 + x2 + x3
update         =~ x4 + x5 + x6
```

```
inhibit        =~ v1*x7 + v1*x8
general        ~~ 0*shift
general        ~~ 0*update
general        ~~ 0*inhibit
shift          ~~ 0*update
shift          ~~ 0*inhibit
update         ~~ 0*inhibit
'

binoinh <- '
general        =~ x1 + x2 + x3  + x4 + x5 + x6 + x7 + x8
shift          =~ x1 + x2 + x3
update         =~ x4 + x5 + x6
general        ~~ 0*shift
general        ~~ 0*update
shift          ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Hull2008.R

```
# Location
folder = 'Hull 2008'
file = 'Hull2008'

# Paper data
N = 100
vars = c(1:10) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                        vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General        =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
'

multimodel <- '
shift          =~ x1 + x2
update         =~ x3 + x4 + x5 + x6 + x7
inhibit        =~ x8 + x9 + x10
'

shiupdmerge <- '
inhibit      =~ x8 + x9 + x10
shiupd       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
'

inhupdmerge <- '
inhupd       =~ x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
shift        =~ x1 + x2
'

inhshimerge <- '
inhshi       =~ x1 + x2 + x8 + x9 + x10
update       =~ x3 + x4 + x5 + x6 + x7
'

bifactor <- '
general        =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
shift          =~ v1*x1 + v1*x2
update         =~ x3 + x4 + x5 + x6 + x7
inhibit        =~ x8 + x9 + x10
```

```
general       ~~ 0*shift
general       ~~ 0*update
general       ~~ 0*inhibit
shift       ~~ 0*update
shift       ~~ 0*inhibit
update       ~~ 0*inhibit
'

binoinh <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
shift       =~ v1*x1 + v1*x2
update       =~ x3 + x4 + x5 + x6 + x7
general       ~~ 0*shift
general       ~~ 0*update
shift       ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Ito2015.R

```
# Location
folder = 'Ito 2015'
file = 'Ito2015'

# Paper data
N = 484
vars = c(1:9) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models

unimodel <- '
General       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
'

multimodel <- '
inhibit     =~ x1 + x2 + x3
update      =~ x4 + x5 + x6
shift       =~ x7 + x8 + x9
'

shiupdmerge <- '
inhibit     =~ x1 + x2 + x3
shiupd      =~ x4 + x5 + x6 + x7 + x8 + x9
'

inhupdmerge <- '
inhupd      =~ x1 + x2 + x3 + x4 + x5 + x6
shift       =~ x7 + x8 + x9
'

inhshimerge <- '
inhshi      =~ x1 + x2 + x3 + x7 + x8 + x9
update      =~ x4 + x5 + x6
'

bifactor <- '
general      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
inhibit      =~ x1 + x2 + x3
update       =~ x4 + x5 + x6
```

```
shift          =~ x7 + x8 + x9
general        ~~ 0*shift
general        ~~ 0*update
general        ~~ 0*inhibit
shift          ~~ 0*update
shift          ~~ 0*inhibit
update         ~~ 0*inhibit
'

binoinh <- '
general        =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
update         =~ x4 + x5 + x6
shift          =~ x7 + x8 + x9
general        ~~ 0*shift
general        ~~ 0*update
shift          ~~ 0*update
'


# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
 runAndSave(sigmaBoot, N, model = binoinh, file = file,
         modelName = 'binoinh')
```

paper_Klauer2010Study1.R

```
# Location
folder = 'Klauer 2010'
file = 'Klauer2010Study1'

# Paper data
N = 125
vars = c(1:7) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)


# Specify Models
unimodel <- '
General       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
'

multimodel <- '
shift        =~ x1 + x2 + x3
inhibit      =~ x4 + x5
update       =~ x6 + x7
'

shiupdmerge <- '
shiupd       =~ x1 + x2 + x3 + x6 + x7
inhibit      =~ x4 + x5
'

inhupdmerge <- '
shift        =~ x1 + x2 + x3
inhupd       =~ x4 + x5 + x6 + x7
'

inhshimerge <- '
update       =~ x6 + x7
inhshi       =~ x1 + x2 + x3 + x4 + x5
'

bifactor <- '
general      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
shift        =~ x1 + x2 + x3
inhibit      =~ v1*x4 + v1*x5
```

```
update        =~ v2*x6 + v2*x7
general       ~~ 0*shift
general       ~~ 0*update
general       ~~ 0*inhibit
shift       ~~ 0*update
shift       ~~ 0*inhibit
update        ~~ 0*inhibit
'

binoinh <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
shift       =~ x1 + x2 + x3
update        =~ v2*x6 + v2*x7
general       ~~ 0*shift
general       ~~ 0*update
shift       ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Klauer2010Study2.R

```
# Location
folder = 'Klauer 2010'
file = 'Klauer2010Study2'

# Paper data
N = 118
vars = c(1:10) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)


# Specify Models
unimodel <- '
General       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
'

multimodel <- '
shift        =~ x1 + x2 + x3
inhibit      =~ x4 + x5 + x6 + x7
update       =~ x8 + x9 + x10
'

shiupdmerge <- '
shiupd       =~ x1 + x2 + x3 + x8 + x9 + x10
inhibit      =~ x4 + x5 + x6 + x7
'

inhupdmerge <- '
shift        =~ x1 + x2 + x3
inhupd       =~ x4 + x5 + x6 + x7 +x8 + x9 + x10
'

inhshimerge <- '
update       =~ x8 + x9 + x10
inhshi       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
'

bifactor <- '
general      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
shift        =~ x1 + x2 + x3
inhibit      =~ x4 + x5 + x6 + x7
```

```
update       =~ x8 + x9 + x10
general      ~~ 0*shift
general      ~~ 0*update
general      ~~ 0*inhibit
shift       ~~ 0*update
shift       ~~ 0*inhibit
update       ~~ 0*inhibit
'

binoinh <- '
general      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
shift       =~ x1 + x2 + x3
update       =~ x8 + x9 + x10
general      ~~ 0*shift
general      ~~ 0*update
shift       ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Lehto2003.R

```
# Location
folder = 'Lehto 2003'
file = 'Lehto2003'

# Paper data
N = 103
vars = c(2,4,7,8,9,10,11,13) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General     =~ x4 + x13 + x7 + x9 + x10 + x11 + x2 + x8
'

multimodel <- '
inhibit     =~ x4 + x13
update      =~ x7 + x9 + x10 + x11
shift       =~ x2 + x8
'

shiupdmerge <- '
inhibit     =~ x4 + x13
shiupd      =~ x7 + x9 + x10 + x11 + x2 + x8
'

inhupdmerge <- '
shift       =~ x2 + x8
inhupd      =~ x4 + x13 + x7 + x9 + x10 + x11
'

inhshimerge <- '
update      =~ x7 + x9 + x10 + x11
inhshi      =~ x4 + x13 + x2 + x8
'

bifactor <- '
general     =~ x4 + x13 + x7 + x9 + x10 + x11 + x2 + x8
inhibit     =~ v1*x4 + v1*x13
update      =~ x7 + x9 + x10 + x11
shift       =~ v2*x2 + v2*x8
```

```
general       ~~ 0*shift
general       ~~ 0*update
general       ~~ 0*inhibit
shift        ~~ 0*update
shift        ~~ 0*inhibit
update       ~~ 0*inhibit
'

binoinh <- '
general        =~ x4 + x13 + x7 + x9 + x10 + x11 + x2 + x8
update         =~ x7 + x9 + x10 + x11
shift         =~ v2*x2 + v2*x8
general       ~~ 0*shift
general       ~~ 0*update
shift        ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Miller2012.R

```
# Location
folder = 'Miller 2012'
file = 'Miller2012'

# Paper data
N = 129
vars = c(1:5,7:9,11:13) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                        vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General        =~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x9 + x11 + x12 + x13
'

multimodel <- '
update         =~ x1 + x2 + x3 + x4
inhibit        =~ x5 + x7 + x8 + x9
shift          =~ x11 + x12 + x13
'

shiupdmerge <- '
shiupd         =~ x1 + x2 + x3 + x4 + x11 + x12 + x13
inhibit        =~ x5 + x7 + x8 + x9
'

inhupdmerge <- '
shift          =~ x11 + x12 + x13
inhupd         =~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x9
'

inhshimerge <- '
update         =~ x1 + x2 + x3 + x4
inhshi         =~ x5 + x7 + x8 + x9 + x11 + x12 + x13
'

bifactor <- '
general        =~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x9 + x11 + x12 + x13
update         =~ x1 + x2 + x3 + x4
inhibit        =~ x5 + x7 + x8 + x9
shift          =~ x11 + x12 + x13
```

```
general        ~~ 0*shift
general        ~~ 0*update
general        ~~ 0*inhibit
shift        ~~ 0*update
shift        ~~ 0*inhibit
update        ~~ 0*inhibit
'

binoinh <- '
general        =~ x1 + x2 + x3 + x4 + x5 + x7 + x8 + x9 + x11 + x12 + x13
update        =~ x1 + x2 + x3 + x4
shift        =~ x11 + x12 + x13
general        ~~ 0*shift
general        ~~ 0*update
shift        ~~ 0*update
'


# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Miyake2000.R

```
# Location
folder = 'Miyake 2000'
file = 'Miyake2000'

# Paper data
N = 137
vars = c(1:9) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9
'

multimodel <- '
shift        =~ x1 + x2 + x3
update       =~ x4 + x5 + x6
inhibit      =~ x7 + x8 + x9
'

shiupdmerge <- '
shiupd       =~ x1 + x2 + x3 + x4 + x5 + x6
inhibit      =~ x7 + x8 + x9
'

inhupdmerge <- '
shift        =~ x1 + x2 + x3
inhupd       =~ x4 + x5 + x6 + x7 + x8 + x9
'

inhshimerge <- '
update       =~ x4 + x5 + x6
inhshi       =~ x7 + x8 + x9 + x1 + x2 + x3
'

bifactor <- '
general      =~ x1 + x2 + x3  + x4 + x5 + x6 + x7 + x8 + x9
shift        =~ x1 + x2 + x3
update       =~ x4 + x5 + x6
inhibit      =~ x7 + x8 + x9
```

```
general       ~~ 0*shift
general       ~~ 0*update
general       ~~ 0*inhibit
shift         ~~ 0*update
shift         ~~ 0*inhibit
update        ~~ 0*inhibit
'

binoinh <- '
general       =~ x1 + x2 + x3  + x4 + x5 + x6 + x7 + x8 + x9
shift         =~ x1 + x2 + x3
update        =~ x4 + x5 + x6
general       ~~ 0*shift
general       ~~ 0*update
shift         ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Rose2012.R

```
# Location
folder = 'Rose 2012'
file = 'Rose2012'

# Paper data
N = 131
vars = c(1:10) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
'

multimodel <- '
update       =~ x1 + x2 + x3 + x4 + x5 + x6
inhibit      =~ x7 + x8
shift        =~ x9 + x10
'

shiupdmerge <- '
shiupd       =~ x1 + x2 + x3 + x4 + x5 + x6 + x9 + x10
inhibit      =~ x7 + x8
'

inhupdmerge <- '
shift        =~ x9 + x10
inhupd       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8
'

inhshimerge <- '
update       =~ x1 + x2 + x3 + x4 + x5 + x6
inhshi       =~ x7 + x8 + x9 + x10
'

bifactor <- '
general      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
update       =~ x1 + x2 + x3 + x4 + x5 + x6
inhibit      =~ v1*x7 + v1*x8
shift        =~ v2*x9 + v2*x10
```

```
general       ~~ 0*shift
general       ~~ 0*update
general       ~~ 0*inhibit
shift        ~~ 0*update
shift        ~~ 0*inhibit
update       ~~ 0*inhibit
'

binoinh <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10
update       =~ x1 + x2 + x3 + x4 + x5 + x6
shift        =~ v2*x9 + v2*x10
general       ~~ 0*shift
general       ~~ 0*update
shift        ~~ 0*update
'


# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Usai2014.R

```
# Location
folder = 'Usai 2014'
file = 'Usai2014'

#Paper data

N = 175
vars = c(1:6) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                        vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General        =~ x1 + x2 + x3 + x4 + x5 + x6
'

multimodel <- '
inhibit        =~ x1 + x2
update         =~ x3 + x4
shift          =~ x5 + x6
'

shiupdmerge <- '
inhibit       =~ x1 + x2
shiupd        =~ x3 + x4 + x5 + x6
'

inhupdmerge <- '
inhupd        =~ x1 + x2  + x3 + x4
shift         =~ x5 + x6
'

inhshimerge <- '
inhshi        =~ x1 + x2 + x5 + x6
update        =~ x3 + x4
'

bifactor <- '
general        =~ x1 + x2 + x3 + x4 + x5 + x6
inhibit        =~ v1*x1 + v1*x2
update         =~ v2*x3 + v2*x4
```

```
shift          =~ v3*x5 + v3*x6
general        ~~ 0*inhibit
general        ~~ 0*shift
general        ~~ 0*update
inhibit        ~~ 0*shift
inhibit        ~~ 0*update
shift          ~~ 0*update
'

binoinh <- '
general        =~ x1 + x2 + x3 + x4 + x5 + x6
update         =~ v2*x3 + v2*x4
shift          =~ v3*x5 + v3*x6
general        ~~ 0*shift
general        ~~ 0*update
shift          ~~ 0*update
'

# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Xu2013Ages7to9.R

```
# Location
folder = 'Xu 2013'
file = 'Xu2013Ages7to9'

# Paper data
N = 140
vars = c(1:7) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                        vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General     =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
'

multimodel <- '
update      =~ x1 + x2 + x3
inhibit     =~ x4 + x5
shift       =~ x6 + x7
'

shiupdmerge <- '
shiupd      =~ x1 + x2 + x3 + x6 + x7
inhibit     =~ x4 + x5
'

inhupdmerge <- '
shift       =~ x6 + x7
inhupd      =~ x1 + x2 + x3 + x4 + x5
'

inhshimerge <- '
update      =~ x1 + x2 + x3
inhshi      =~ x4 + x5 + x6 + x7
'

bifactor <- '
general     =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
update      =~ x1 + x2 + x3
inhibit     =~ v1*x4 + v1*x5
shift       =~ v2*x6 + v2*x7
```

```
general       ~~ 0*shift
general       ~~ 0*update
general       ~~ 0*inhibit
shift        ~~ 0*update
shift        ~~ 0*inhibit
update       ~~ 0*inhibit
'

binoinh <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
update        =~ x1 + x2 + x3
shift        =~ v2*x6 + v2*x7
general       ~~ 0*shift
general       ~~ 0*update
shift        ~~ 0*update
'


# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Xu2013Ages10to12.R

```
# Location
folder = 'Xu 2013'
file = 'Xu2013Ages10to12'

# Paper data
N = 165
vars = c(1:7) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
'

multimodel <- '
update       =~ x1 + x2 + x3
inhibit      =~ x4 + x5
shift        =~ x6 + x7
'

shiupdmerge <- '
shiupd       =~ x1 + x2 + x3 + x6 + x7
inhibit      =~ x4 + x5
'

inhupdmerge <- '
shift        =~ x6 + x7
inhupd       =~ x1 + x2 + x3 + x4 + x5
'

inhshimerge <- '
update       =~ x1 + x2 + x3
inhshi       =~ x4 + x5 + x6 + x7
'

bifactor <- '
general      =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
update       =~ x1 + x2 + x3
inhibit      =~ v1*x4 + v1*x5
shift        =~ v2*x6 + v2*x7
```

```
general       ~~ 0*shift
general       ~~ 0*update
general       ~~ 0*inhibit
shift       ~~ 0*update
shift       ~~ 0*inhibit
update       ~~ 0*inhibit
'

binoinh <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
update       =~ x1 + x2 + x3
shift       =~ v2*x6 + v2*x7
general       ~~ 0*shift
general       ~~ 0*update
shift       ~~ 0*update
'


# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

paper_Xu2013Ages13to15.R

```
# Location
folder = 'Xu 2013'
file = 'Xu2013Ages13to15'

# Paper data
N = 152
vars = c(1:7) # Which rows/columns of cor. matrix to import
Sigma = importCorMatrix(paste(folder, '/', file, '.txt', sep = ''),
                    vars = vars)

# Bootstrap correlation matrices
sigmaBoot = bootstrapCor(Sigma, N, samples)

# Specify Models
unimodel <- '
General       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
'

multimodel <- '
update        =~ x1 + x2 + x3
inhibit       =~ x4 + x5
shift         =~ x6 + x7
'

shiupdmerge <- '
shiupd        =~ x1 + x2 + x3 + x6 + x7
inhibit       =~ x4 + x5
'

inhupdmerge <- '
shift         =~ x6 + x7
inhupd        =~ x1 + x2 + x3 + x4 + x5
'

inhshimerge <- '
update        =~ x1 + x2 + x3
inhshi        =~ x4 + x5 + x6 + x7
'

bifactor <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
update        =~ x1 + x2 + x3
inhibit       =~ v1*x4 + v1*x5
shift         =~ v2*x6 + v2*x7
```

```
general       ~~ 0*shift
general       ~~ 0*update
general       ~~ 0*inhibit
shift         ~~ 0*update
shift         ~~ 0*inhibit
update        ~~ 0*inhibit
'

binoinh <- '
general       =~ x1 + x2 + x3 + x4 + x5 + x6 + x7
update        =~ x1 + x2 + x3
shift         =~ v2*x6 + v2*x7
general       ~~ 0*shift
general       ~~ 0*update
shift         ~~ 0*update
'


# (1) Fit unidimensional model
runAndSave(sigmaBoot, N, model = unimodel, file = file,
        modelName = 'unimodel')

# (2) Fit three factor model
runAndSave(sigmaBoot, N, model = multimodel, file = file,
        modelName = 'multimodel')

# (3) Fit shift-update merged model
runAndSave(sigmaBoot, N, model = shiupdmerge, file = file,
        modelName = 'shiupdmerge')

# (4) Fit inhibition-update merged model
runAndSave(sigmaBoot, N, model = inhupdmerge, file = file,
        modelName = 'inhupdmerge')

# (5) Fit inhibition-shift merged model
runAndSave(sigmaBoot, N, model = inhshimerge, file = file,
        modelName = 'inhshimerge')

# (6) Fit bifactor model
runAndSave(sigmaBoot, N, model = bifactor, file = file,
        modelName = 'bifactor')

# (7) Fit bifactor model with no inhibition
runAndSave(sigmaBoot, N, model = binoinh, file = file,
        modelName = 'binoinh')
```

Simulation Code

Provided on the following pages is a set of R code that was used to run the simulation. Each page represents a separate R file. The first four pages (*bootstrapCor.R*; *importCorMatrix.R*; *modelFit.R*; and *runAndSave.R*) were included in the Sim_code subfile as per the directory structure specified earlier. The file page (*run_Simulation.R*) was included in the R_Code_Simulation superfolder, and was not placed in any subfolder.

bootstrapCor.R

```
bootstrapCor = function(Sigma, N, samples){


   storage = array(NA, dim = c(dim(Sigma), samples),
                dimnames = list(row.names(Sigma), NULL, NULL))

   for (i in 1:samples){
      x = data.frame(mvrnorm(n = N,
                     mu = rep(0, dim(Sigma)[1]),
                     Sigma = Sigma))
      storage[,,i] = cor(x)
   }

   return(storage)
}
```

importCorMatrix.R

```
importCorMatrix = function(file, vars){

  # Import datafile
  file = paste('paper_data/', file, sep = '')
  dataFile = read.table(file, sep = '', header = F, fill = TRUE)

  # Construct covariance matrix
  corMatrix = as.matrix(dataFile[-c(1,2),])
  corMatrix = matrix(mapply(FUN = sum,
                            corMatrix,
                            t(corMatrix),
                            -diag(rep(1, dim(corMatrix)[1])),
                            MoreArgs = list(na.rm = TRUE)),
                     nrow = dim(corMatrix)[1])
  corMatrix = corMatrix[vars, vars]
  row.names(corMatrix) = paste('x', vars, sep = '')

  return(corMatrix)
}
```

modelFit.R

```
modelFit = function(Sigma, N, model, sigmaBoot){
  fit = cfa(model,
            sample.cov = Sigma,
            sample.mean = rep(0, dim(Sigma)[1]),
            sample.nobs = N,
            mimic = 'Mplus',
            std.lv = T)

  # Cross-validation was calculcated but was not evaluated
  # Mean of RMSEs across all bootstrap iterations
  pred <- fitted(fit)$cov
  D = 1/sqrt(diag(pred))
  pred = D * pred * D
  pred <- pred[lower.tri(pred)]
  cv <- mean(apply(sigmaBoot, 3, function(x) sqrt(mean((x[lower.tri(x)] - pred)^2))  ))
  names(cv) <- 'CV'

  # Get factor correlations
  corNames <- names(coef(fit))
  corIdx   <- which(corNames %in% c('update~~shift',   'shift~~update',
                                   'update~~inhibit', 'inhibit~~update',
                                   'inhibit~~shift',  'shift~~inhibit',
                                   'inhibit~~shiupd', 'shiupd~~inhibit',
                                   'shift~~inhupd',   'inhupd~~shift',
                                   'update~~inhshi',  'inhshi~~update'))
  cors     <- coef(fit)[corIdx]

   #omega reliability was also calculated but not evalauted

  return(c(reliability(fit)['omega3',],
        fitMeasures(fit)[c(5,9,19,20,22,23)], cv, cors))
}
```

runAndSave.R

```
runAndSave <- function(sigmaBoot, N, model, file, modelName){

   # Apply model to each bootstrap covariance matrix
   fit = alply(sigmaBoot, 3,
            function (x) tryCatch(modelFit(x, N, model, sigmaBoot),
                            error = function(e) NA,
                            warning = function(w) NA))

   # Convert result to data frame
   frame = data.frame(t(as.data.frame(fit)))

   # Save data file
   write.table(frame,
            file = paste('simulation_data/', file, '_', modelName,
                  '.csv', sep = ''),
            sep = ',', row.names = F)

   return(NULL)
}
```

run_Simulation.R

```
# Sim functions
source('sim_code/modelFit.R')
source('sim_code/importCorMatrix.R')
source('sim_code/bootstrapCor.R')
source('sim_code/runAndSave.R')


# Packages
library(lavaan)
library(semTools)
library(MASS)
library(plyr)

# Parameters
samples = 5000
files = list.files('papers', full.names = T)
for (i in 1:length(files)) source(files[i])
```