Psychological Assessment
2011, Vol. 23, No. 1, 64–79

1040-3590/11/$12.00 DOI: 10.1037/a0021097

Theoretical and Statistical Derivation of a Screener for the Behavioral Assessment of Executive Functions in Children

Mauricio A. Garcia-Barrera, Randy W. Kamphaus, and Deborah Bandalos
University of Georgia

The problem of valid measurement of psychological constructs remains an impediment to scientific progress, and the measurement of executive functions is not an exception. This study examined the statistical and theoretical derivation of a behavioral screener for the estimation of executive functions in children from the well-established Behavior Assessment System for Children (BASC). The original national standardization sample of the BASC–Teacher Rating Scales for children ages 6 through 11 was used ($N = 2{,}165$). Moderate-to-high internal consistency was obtained within each factor (.80–.89). A panel of experts was used for content validity examination. A confirmatory factor analysis model with 25 items loading on 4 latent factors (behavioral control, emotional control, attentional control, and problem solving) was developed, and its statistical properties were examined. The multidimensional model demonstrated adequate fit, and it was deemed invariant after configural, metric, and scalar measurement invariance tests across sex and age. Given its strong psychometric properties, with further tests of item validity, this instrument promises future clinical and research utility for the screening of executive functions in school-age children.

Keywords: screening, executive functions, instrument development, behavioral assessment, rating scale

As awareness of the importance of executive functions grows, several efforts have been made to improve both its theoretical definition and the sophistication of measurement techniques to capture this complex function. Discussion about how to define executive functioning has unfolded into a series of debates about its configuration (i.e., unitary vs. diverse), its nature (i.e., biological or theoretical), and its composition (i.e., underlying components). In this plurality of debates, there are almost as many definitions of executive functions as there are researchers using the construct. The burst of newly developed instruments based in somewhat new models is a good example of this plurality. The most parsimonious definition of executive function includes the idea of an organized system of abilities working toward the conception, monitoring, and execution of goal-directed behavior (Ardila, 2008; Jurado & Roselli, 2007; Royall et al., 2002). In reality, and as Elliot (2003) asserted, "there is no intuitive lay concept that incorporates the essence of executive function" (p. 50).

Recent work by Miyake et al. (2000) has demonstrated the utility of using sophisticated statistical modeling in the attempt to develop psychometrically derived or evidence-based models of executive functions. In this regard, this study presents a piloting statistical derivation of a behavioral screener based on four out of the five components of executive functioning proposed by Garcia-Barrera (2010). The original model includes a goal identification component labeled *problem solving*, an updating working memory component (not evaluated in this study), and three commonly identified cybernetic aspects of executive functioning: attentional control, behavioral control, and emotional control. The four components included in this study are briefly defined as follows.

Executive Functioning: Theoretical Model

Problem Solving

The problem-solving construct is associated with goal identification and the subsequent initiation of behavior. Luria (1973) described problem solving in terms of the creation of intentions, plans, and programs; Lezak, Howieson, and Loring's (2004) model defined it in terms of volition, planning, and purposive action; Fuster (2008) defined it in terms of the temporal organization of behavior; Royall et al. (2002) called the problem-solving functions *executive cognitive functions*; Denckla (2007) referred to problem solving as the *how* executive functions. In the model we propose,

This article was published Online First January 17, 2011.

Mauricio A. Garcia-Barrera, Randy W. Kamphaus, and Deborah Bandalos, Educational Psychology and Instructional Technology, University of Georgia. Mauricio A. Garcia-Barrera is now at the Department of Psychology, University of Victoria, Victoria, British Columbia, Canada. Randy W. Kamphaus is now at the Department of Counseling and Psychological Services, Georgia State University. Deborah Bandalos is now at the Department of Graduate Psychology, James Madison University.

This research was conducted as the doctoral dissertation of Mauricio A. Garcia-Barrera and was supported in part by a University of Georgia Graduate School Dean's Research Award. Preliminary analyses were presented at the 35th (Portland, Oregon; February 2007) and 36th (Waikoloa, Hawaii; February 2008) International Neuropsychological Society annual meetings. Randy W. Kamphaus is a coauthor of the Behavior Assessment System for Children, and as such he receives royalties on sales. We acknowledge the valuable contributions of George W. Hynd, Jonathan Campbell, L. Stephen Miller, and Morris Cohen to the study conceptualization.

Correspondence concerning this article should be addressed to Mauricio A. Garcia-Barrera, Department of Psychology, University of Victoria, P.O. Box 3050 STN SCS, Victoria, British Columbia, Canada V8W 3P5. E-mail: mgarcia@uvic.ca

problem solving is an umbrella construct incorporating abilities used to create a response to a novel situation, or *preparation to action*; at the behavioral level, this is associated with planning, making decisions, conflict resolution, and organizing information toward the execution of a goal. Zelazo, Carter, Reznick, and Frye (1997) presented a problem-solving model of executive functions that emphasizes problem representation, planning, execution, and subsequent evaluation as crucial executive function components. Some traditional information-processing models of executive function have referred to this construct as *task analysis* (Borkowski & Burke, 1996).

From a neuroanatomical perspective, Fuster (2008) proposed that, because of its location within the frontal lobe, the prefrontal cortex has a specific role in the temporal organization of behavior. Problem solving was a factor included to attempt to indirectly (i.e., behaviorally) measure how effectively behavior is organized and programmed toward goal attainment. Within the prefrontal cortex, the dorsolateral prefrontal circuit has been involved in planning, self-monitoring, and other higher cognitive functions (Royall et al., 2002).

Attentional Control

This construct is related to the ability to focus, sustain, and shift attentional systems according to task demands. These three elements of attention have been derived from Mirsky, Anthony, Duncan, Ahearn, and Kellam's (1991) model but under the assumption that executive functions do have a regulatory role over these attentional systems, as was presented earlier by Norman and Shallice's (1986) Supervisory Attentional System (SAS; Shallice & Burguess, 1996) and was further supported with the work of Baddeley (1996) on the central executive and the research of Fuster (2002, 2008, 2008) on prefrontal functioning. Furthermore, Posner and colleagues (Posner & Rothbart, 1998; Rueda, Posner, & Rothbart, 2005) have asserted that voluntary, controlled attention is associated with dopaminergic-modulated networks, such as the anterior cingulate, basal ganglia, and lateral prefrontal cortex, which are responsible for the executive attention, or executive control, system. The proposed model assumes Posner and colleagues' executive attention to be well represented under the attentional control component of the executive function model.

Behavioral Control

This construct was included in the model under the assumption that executive functions play an important role in the self-regulation of behavior, including inhibition/impulse control (Garcia-Barrera, 2010). As Barkley (1997) asserted, behavioral inhibition refers to the ability to inhibit initial prepotent responses to external cues or events, to stop an ongoing response producing a delay period, and to protect this delay from interferences (e.g., competing events). Self-control refers to the responses by the individual that "serve to alter the probability of their subsequent response to an event" (Barkley, 1997, p. 51), also influencing the probabilities of occurrence of the consequences naturally to follow. Behavioral self-regulation has also been associated with internal self-directed speech. The anterior cingulate cortex and the lateral orbitofrontal areas appear to be involved in the initiation of behaviors and the inhibition of prepotent behavioral responses (Rolls, 2002; Royall et al., 2002).

Emotional Control

This construct represents the ability to self-regulate emotional response to environmental and internal cues (Garcia-Barrera, 2010). Emotional self-regulation, or control, is highly related to behavioral inhibition and self control. However, it is differentiated by the fact that the prepotent response that is being delayed is the expression of emotional reactions "that would have been elicited by the event and whose expression would have been a part of the expression of those prepotent responses" (Barkley, 1997, p. 182). The need for inclusion of the role of emotions in decision making was originally proposed by Damasio (1995). Furthermore, emotional regulation facilitates the efficiency of the system to read environmental cues and optimizes recall of information from long-term memory into working memory by means of resource allocation.

Altogether, attentional control, behavioral control, and emotional control represent the *cybernetic* aspect of executive function (Royall et al., 2002), encompassing some of the potential relationships that executive systems have with nonexecutive systems. Moreover, these three elements could be grouped under the *when* executive functions as presented by Denckla (2007).

Finally, there is one more component in Garcia-Barrera's (2010) model: updating working memory representations. Although it was not included in the screener derivation presented here, the underlying assumption is that during the process of generating plans and programs for effective problem solving, as well as during the activation and maintenance of self-regulatory attentional, behavioral, and emotional systems, the information needs to be processed and manipulated in an online system: working memory. Baddeley's (1996) central executive component in working memory illustrates this component. In the model presented by Garcia-Barrera (2010), the continuous and timely capacity to update working memory representations according to task demands was specifically included as a component of the executive system (Lehto, 1996; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Miyake, Emerson, & Friedman, 2000; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001).

Measuring Executive Functions: Behavioral and Ecological Approach

Early models in neuropsychology considered the Wisconsin Card Sorting Test, the Towers of Hanoi and London and other variations, and even the Rey-Osterrieth Complex Figure as examples of gold-standard tools for assessing executive function (Anderson, 2001; Lezak, Howieson, & Loring, 2004). However, criticisms of the limitations on sensitivity and specificity of these and other instruments created the need for a more multidimensional approach (Anderson, 2001; Hughes & Graham, 2002; Manchester, Priestley, & Jackson, 2004). In addition some measures used to collect information about executive functioning may be lengthy or require apparatus, making them impractical for widespread use. Lengthy instruments, in particular, may be impractical for the respondent (Flanagan, Bierman, & Kam, 2003). Practicality is important because practical instruments may be used on a larger

scale, thus allowing more children with executive functions problems to be identified, which in turn mitigates functional impairment (Kamphaus et al., 2007). Thus, the assessment of child executive functions needs both psychometric and practical improvements.

Recently, a psychometric approach developed from the raising awareness of ecological validity of assessment instruments, and it has been focused on the analysis of the everyday behavioral components of executive functions. This approach gained recognition with the development of behavioral ratings of frontal and executive function. Examples from the adult assessment literature include the Frontal Systems Behavior Scales (Grace & Malloy, 2001) and the rating included on the Dysexecutive Questionnaire–Behavioural Assessment of the Dysexecutive Syndrome (Wilson, Evans, Emslie, Alderman, & Burgess, 1998). There are very few options for the behavioral assessment of children's executive functioning; the most popular measure is the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000).

However, average and unimpaired executive systems at different normal developmental stages were overlooked for decades, and research on the assessment of executive function was heartily criticized for neglecting to include analysis of average children populations. In fact, it was only recently that a burst of studies of executive functions in preschoolers and school-age children was published (Carlson, 2005; Espy, 2004; Espy, Kaufmann, & Glisky, 2001; Hughes & Graham, 2002; Isquith, Gioia, & Espy, 2004; Senn, Espy, & Kaufmann, 2004). Developmental psychologists have made considerable contributions to the understanding of the development of executive functions from early childhood to school age (e.g., Zelazo, Müeller, Frye, & Marcovitch, 2003), the clinical impairment of executive functions (e.g., Hughes, 2002), and the relationship between brain development and the acquisition of executive skills (e.g., Diamond, 2002; Tsujimoto, 2008). Furthermore, executive functions, such as problem solving, behavioral control, attentional control, and emotional control, are important in students' everyday routines (Clark, Prior, & Kinsella, 2002), and the development of more reliable and valid instruments to estimate executive function in children appears imperative.

Screening measures are a potential avenue for the early assessment and identification of executive dysfunctions during first layer assessments in school and pediatric settings. Behavioral screeners are effective, efficient, and costless measures (Lochman, 1995). They can be developed around parsimonious models that allow replicability and cross-cultural utility, and they can provide fundamental information for neuroanatomical-correlation studies. Besides the BRIEF, we are unaware of other short screening measures available for the behavioral estimation of executive functions in school-age children; yet, we recognize the potential of some existing behavioral tools to aid clinicians in estimating executive-like behaviors. In this regard, the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992) is a multidimensional rating scale of externalizing, internalizing, and adaptive skills that includes questionnaires for parents and teachers of children and adolescents between the ages of 2 and 18. The BASC is among the most widely used measures of child behavior in the United States (Reynolds & Kamphaus, 2002). It has been cross-culturally validated in Colombia by Pineda et al. (1999), and it has been broadly validated and published in Spain, among other countries (http://www.teaediciones.com).

The validity of the BASC as an assessment tool for frontal lobe/executive function has been previously studied (Jarratt, Riccio, & Siekierski, 2005; Mahone, Zabel, Levey, Verda, & Kinsman, 2002; Riccio et al., 1994). However, only one recent study utilized the Frontal Lobe Functioning/Executive Control scale recently included in the BASC-2 extended software (Sullivan & Riccio, 2006). Sullivan and Riccio (2006) administered the original 18-item BASC Frontal Lobe Functioning/Executive Control scale, the BRIEF, and the Conners' Parent Rating Scales Revised–Short Form (Connors, 1997) to a community sample of 92 children with or without attention-deficit/hyperactivity disorder (ADHD). This 18-item scale was originally developed by Barringer and Reynolds (1995) and was presented as a paper at the annual meeting of the National Academy of Neuropsychology. Sullivan and Riccio's study found the BASC Frontal Lobe Functioning/Executive Control scale to be sensitive to the identification of behaviors associated with executive dysfunctions in children with ADHD and other disorders, and significant correlations with scores on all the scales of the BRIEF and Conners' Parent Rating Scales Revised–Short Form were reported. Correlations with the BRIEF scales ranged from .45 (Organization of Materials) to .83 (Global Executive Composite), and correlations with the Conners' Parent Rating Scales Revised–Short Form scales ranged from .63 (ADHD index) to .77 (Oppositional scale). According to Sullivan and Riccio's report (p. 499), all correlations were significant at the $p < .001$ level.

Because of the identified need for a screening measure of executive functions in the assessment of school-age children, this study aimed to augment early work by Barringer and Reynolds (1995) and statistically derive a screening instrument under two constraints: First, to avoid proliferation of more testing measurements, we derived the screener from a well-known, established, reliable, and valid existing measure: the BASC–Teacher Rating Scale for children ages 6 through 11. This way, one administration would provide the examiner with both a comprehensive evaluation of a child's behavior and an estimation of executive behavior. This screener comprises 25 items and is based on a theoretically driven multidimensional model (Garcia-Barrera, 2010) with four factors representing four core executive functions: attentional control, behavioral control, emotional control, and problem solving. Second, we tested this screener's underlying four-factor model with rigor, using structural equation modeling and guided by the following question: Is it possible to derived a statistically and theoretically based measure of children's executive functions from an existing behavioral rating scale?

Finally, because of the promising findings presented by Barringer and Reynolds (1995) and by Sullivan and Riccio (2006), a stringent item selection and statistical examination, and a sound theoretical model of executive function, we hypothesized that the screener derivation would succeed in identifying a set of items measuring those four core areas (i.e., problem solving and attentional, emotional, and behavioral control) while holding the statistical properties of the model underlying it (e.g., factor internal consistency levels $> .80$ and model fit indexes $> .90$).

Methods

The BASC–Teacher Rating Scale–Child (BASC-TRS-C; Reynolds & Kamphaus, 1992), which is used to rate children ages 6 through 11, was selected as the target instrument for the screener derivation to complement prior derivation attempts that used the BASC–Parent Rating Scale (Barringer & Reynolds, 1995; Sullivan & Riccio, 2006) and under the assumption that teachers are more accurate raters than parents of children's cognitive abilities (e.g., executive functions) and behavior (Lochman, 1995; Reynolds & Kamphaus, 1992). The BASC-TRS-C yields nine problem behavior scales and four adaptive skills scales as well as composite scores, and it includes 148 items to be rated on a 4-point Likert scale anchored by 1 (*never*) and 4 (*almost always*; Reynolds & Kamphaus, 1992).

Participants and Data Collection

Pearson Assessment gave us permission to utilize the original BASC-TRS-C standardization database. The standardization of the BASC included a vast sample representative of the population of U.S. children, and it included a range of diversity in terms of ethnicity, socioeconomic status, geographic region, and clinical problems (Reynolds & Kamphaus, 1992). Detailed information about the data collection procedures is found in the original BASC manual. Table 1 reports a summary of the demographic characteristics of the sample.

Table 1
Behavior Assessment System for Children–Teacher Rating Scale–Child (6–11 Years) Original Standardization Sample's Demographic Distribution

Variable	%
Demographic representation of the general-norm sample	
Gender	
Male	40
Female	60
Race	
African American	24
Hispanic	7
White	67
Other	2
Geographic region	
Northeast	8
North central	35
South	36
West	21
General norm sample by special education classification and gender	
Behavioral/emotional disturbance	
Female	0.3
Male	0.9
Learning disability	
Female	2.6
Male	5.9
Mild mental retardation	
Female	0.8
Male	1.2
Speech/language disorder	
Female	1.6
Male	1.9
Representation of the clinical norm sample by primary diagnosis or classification	
Behavior disorder	47
Conduct disorder	4
Attention-deficit/hyperactivity disorder	17
Depression	3
Autism	3
Emotional disturbance	3
Undifferentiated	16
Other	6
Representation of the aggregated sample by developmental age	
6 years old	11
7 years old	16
8 years old	17
9 years old	23
10 years old	17
11 years old	16

Note. Adapted from *Behavior Assessment System for Children* (*BASC*), by C. R. Reynolds and R. W. Kamphaus, 1992, pp. 87–93. Copyright 1992 by American Guidance Services.

Screener Derivation Process

Figure 1 delineates the succession of steps (phases) followed during this instrument derivation.

Phase 1: Item selection pool. The 148 items from the original BASC-TRS-C were carefully reviewed to identify items that assess behaviors that can be potentially classified as executive. We extracted an initial set of 28 items. The items themselves are only indicators of the construct, which is latent in nature. We hypothesized that these BASC-TRS-C items could be grouped into four factors that represented four essential executive functions: Problem Solving, Attentional Control, Behavioral Control, and Emotional Control.

Phase 2: Latent construct operationalization and item distribution. Table 2 includes the original BASC-TRS-C items, their scale membership, and their matching items from the BASC-2. (Items from our screener that are also included in the BASC-2 Frontal Lobe Functioning scale are marked to facilitate identification.)

Problem solving. The problem-solving construct measures the ability to plan, problem solve, make decisions, and organize information toward the execution of a goal. It includes 10 items, such as *analyzes the nature of a problem before starting to solve it* and *makes decisions easily*. Ratings in this set of items assess the following question: Does this child demonstrate knowledge of how to achieve a goal?

Attentional control. The attentional control construct includes seven items and measures the ability to focus (e.g., *makes careless mistakes*), sustain (e.g., *has trouble concentrating*), and shift attention systems according to task demands (e.g., *has trouble shifting gears from one task to another*). One memory/working memory item was included (*forgets things*) to include an estimation of the effects of attention problems in memory and therefore in learning. Ratings within this construct attempt to answer the following question: Can this child self-regulate attention?

Behavioral control. The behavioral control construct comprises seven items and was included in this screener under the assumption that executive functions play an important role in the ability to self-regulate behavior (e.g., *acts without thinking*), in-

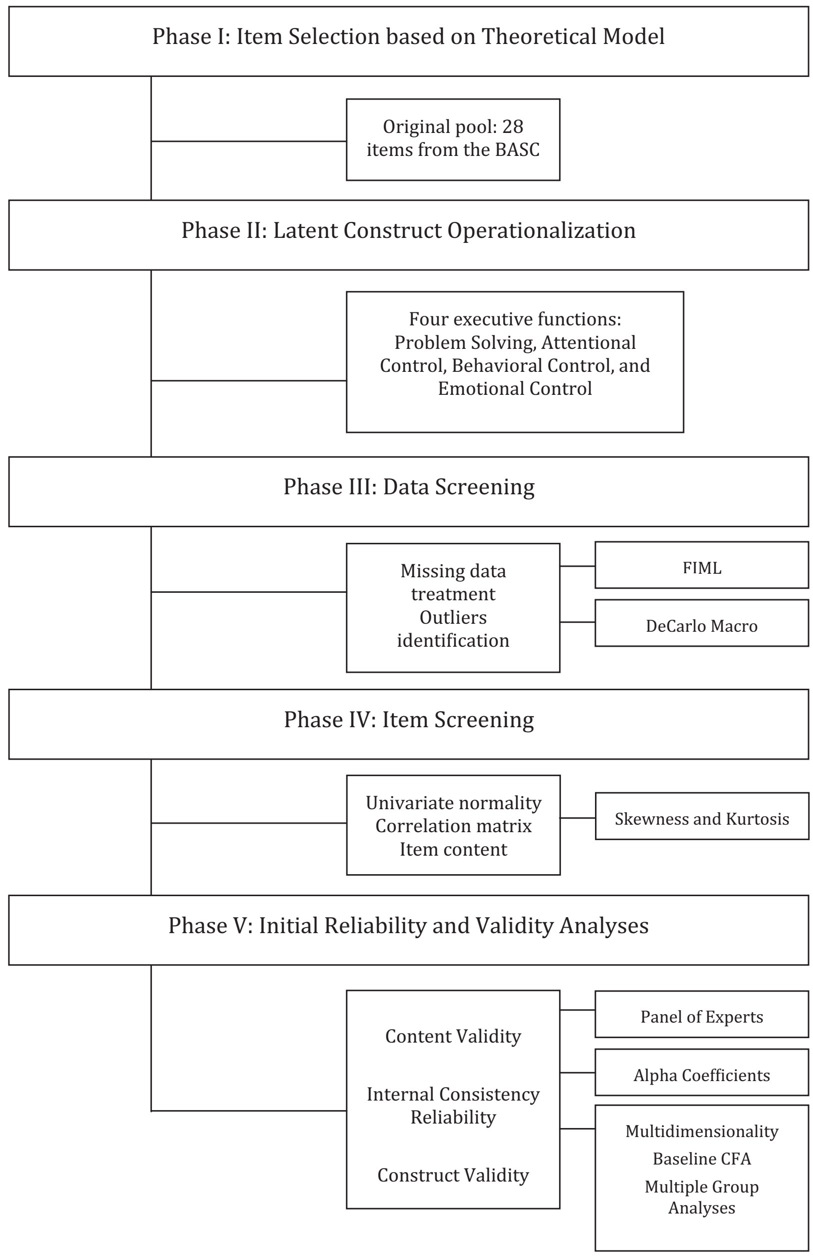

Figure 1. Screener derivation flowchart. BASC = Behavior Assessment System for Children; FIML = full information maximum likelihood; CFA = confirmatory factor analysis.

cluding inhibition (e.g., *uses foul language*) and impulse control (e.g., *interrupts others when they are speaking*). This construct answers the following question: Can this child inhibit prepotent response and self-regulate behavior?

Emotional control. Items under the emotional control construct measure the ability to self-regulate emotional response to environmental and internal cues. This ability is of great importance when trying to examine the relationships between prefrontal areas and the limbic system, in terms of the regulatory function of executive systems over emotion. This factor comprises four items, representing a limited yet significant range of emotional disregulation indicators (e.g., *changes moods quickly* and *throws tantrums*). This construct answers the following question: Can this child self-regulate emotional expression?

Phase 3: Data screening.

Missing data treatment. Typical missing data techniques include listwise and pairwise deletion, mean imputation, regression-based imputation, and hot-deck imputation (Enders & Bandalos, 2001). The validity of such data imputation techniques has been questioned because of their tendency to produce biased parameter estimates. Recently, a full information maximum likelihood approach has been presented as an unbiased missing data

Table 2
Original Distribution of Items Per Scale

BASC	BASC-2	Original scale membership
	Problem Solving	
Item 37	Item 17	Study Skills
Item 50		Study Skills
Item 60	Item 58	Leadership
Item 64		Study skills
Item 83	Item 117	Leadership
Item 110	Item 88	Social Skills
Item 111	Item 129	Study Skills
Item 120	Item 30	Leadership
Item 124	Item 73	Study Skills
Item 148	Item 91	Study Skills
	Attentional Control	
Item 10	Item 20	Learning Problems
Item 18	Item 61[a]/Item 100	Attention Problems
Item 24		Learning Problems
Item 55	Item 5	Attention Problems
Item 78		Attention Problems
Item 115		Attention Problems
Item 126		Adaptability
	Behavioral Control	
Item 2	Item 8[a]	Aggression
Item 39		Aggression
Item 59	Item 46[a]	Hyperactivity
Item 82	Item 54[a]	Hyperactivity
Item 94		Conduct Problems
Item 127	Item 120[a]	Aggression
Item 144	Item 74[a]	Hyperactivity
	Emotional Control	
Item 7		Depression
Item 44		Depression
Item 95	Item 49[a]	Depression
Item 108		None

Note. BASC = Behavior Assessment System for Children.
[a] Included in the Frontal Lobe/Executive Functioning Scale from the Teacher Rating Scale–Child for children ages 6 through 11, BASC-2.

treatment technique for structural equation modeling analysis. According to Enders and Bandalos (2001), this method does not impute missing values but rather estimates parameters and standard errors by "borrowing information from the observed portion of data" (p. 434) and by computing a "casewise likelihood function using only those variables that are observed for [the] case" with missing data (p. 434). Enders and Bandalos performed a Monte Carlo simulation study that demonstrated the relative unbiased and efficient performance of full information maximum likelihood with respect to other missing data techniques. This approach was also superior in proportion of convergence, lower model rejection rates, and lower risk for Type 1 error. Thus, full information maximum likelihood was used to treat missing data in this study.

Treatment of outliers. Tests for multivariate normality and identification of significant outliers were performed with a macro developed by DeCarlo (1997).

Phase 4: Item screening.

Analysis of frequency distributions. We analyzed frequency distributions, including analysis of means, standard deviations, skewness, and kurtosis. For this study, a critical value of $|2.0|$ was used to test skewness (Crocker & Algina, 1986). There is some debate about the adequate critical value to determine significant kurtosis. The BASC items use a 4-point Likert scale. It has been suggested that scales, such as Likert scales, dealing with ordinal (noncontinuous) data often present a nonnormal distribution, because of the nature of the items' content. For this study, values outside the range of $|7.0|$ are considered to be problematic and represent considerable departures from normality, because of the risk of bias that univariate nonnormality presents (Finney & DiStefano, 2006).

Generation of a correlation matrix. To analyze the relationships among items and potential collinearity effects, we generated a correlation matrix. If correlations are too high, collinearity is suspected (Kline, 2005). However, because all the items in the scale are hypothetically expected to measure different aspects of the same construct, moderate correlation values are expected.

Examination of item content and wording. In this case, items that were suspected of being too ambiguous were eliminated.

Phase 5: Initial reliability and validity analysis.

Content validity. A panel of 10 experts in neuropsychology was first informed about the purposes and goals of the screener. Subsequently, they were asked to classify the 28 items into the four hypothesized executive components, in terms of their accuracy to measure problem solving, attentional control, behavioral control, and emotional control. The items were alphabetically arranged (by the first word) to keep the raters unaware of the original classification. A cutoff of 70% interrater agreement was established as a criterion. As a result of the panel consultation, three items were dropped from the screener and one item-factor association within the model was repositioned.

Internal consistency reliability. Coefficient alpha for each group of items (factors) was used for estimating internal consistency reliability (Crocker & Algina, 1986), and our criterion was established as Cronbach's alpha for each scale greater than .8.

Construct validity. For this study, two well-supported methods of construct validation were used: confirmatory factor analysis (CFA) and measurement invariance (or multiple group analysis). For this purpose, a hypothetical model with four factors and 25 indicators was designed, and its statistical properties were tested (see Figure 2). Unnested variations of the hypothetical CFA model were also evaluated including one factor (unidimensional model) or more factors (multidimensional models).

Estimation Method

Mplus 4.0 (L. K. Muthén & Muthén, 2006) was used to perform the CFA. Mplus utilizes a maximum likelihood method of estimation by default when data is continuous. This study uses Likert-type items. To enhance validity, we analyzed data as categorical (ordinal) rather than continuous (Flora & Curran, 2004; B. O. Muthén & Asparouhov, 2002). The Mplus default estimator for categorical data analyses is weighted least squares with mean and variance adjusted (WLSMV). The WLSMV estimator serves as a correction that is less computationally demanding than other options, such as weighted least squares (WLS), and produces estimates that are unbiased, consistent, and efficient. In contrast to other types of estimators (e.g., WLS), WLSMV uses the diagonal of the weight matrix in the estimation instead of the full weight matrix. Like WLS, WLSMV uses the full weight matrix to com-

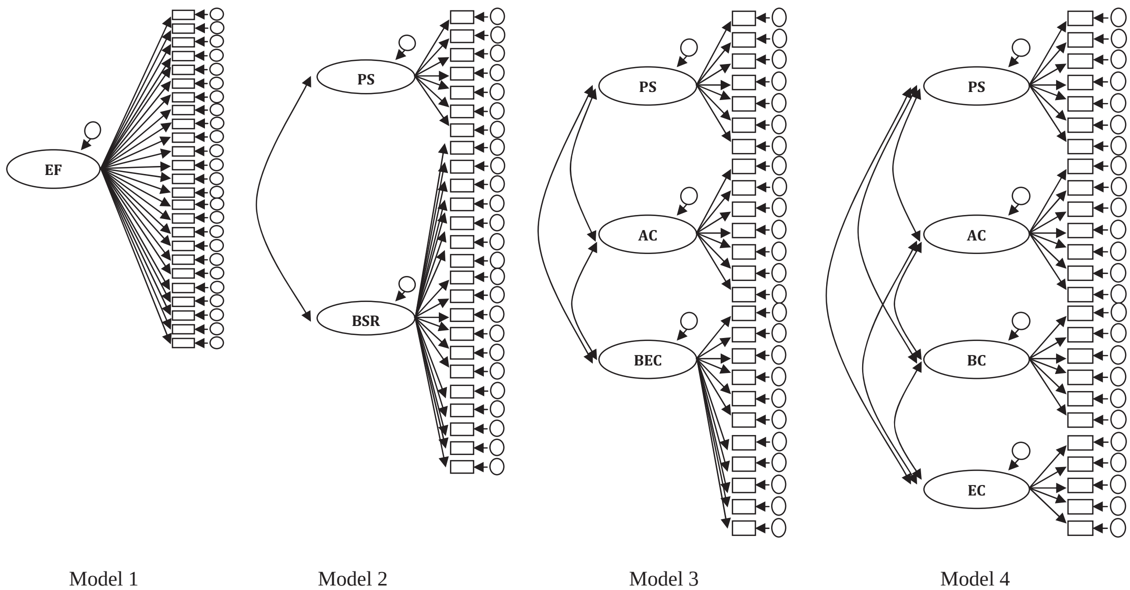

Figure 2. Unidimensional and multidimensional models tested. EF = Executive Functions; PS = Problem Solving; BSR = Behavioral Self-Regulation; AC = Attentional Control; BEC = Behavioral and Emotional Control; BC = Behavioral Control; EC = Emotional Control.

pute standard errors and chi-square (L. K. Muthén & Muthén, 2006). There are two parameterizations provided by Mplus when using WLSMV as the parameter estimator: the conditional probability and the latent response variable formulations. The latent response variable assumes that a continuous and latent response variable (y_i^*) underlies the observed categorical variable (y_i); y_i is related to the latent response variable y_i^* through threshold parameter (T, or intercepts) between categories. For this purpose, polychoric correlations are calculated and used instead of Pearson's correlations (Flora & Curran, 2004).

Fit Indexes

Two types of indicators of overall fit are recommended by Hoyle and Panter (1995) and by Hu and Bentler (1999): the absolute fit and the incremental fit. The absolute fit (or badness of fit) concerns the degree to which the covariances implied by the fixed and free parameters—and specified in the model—match the observed covariances from which free parameters in the model were estimated. Optimal fit is indicated by a value of zero. Hoyle and Panter recommended chi-square accompanied by degree of freedom, sample size, and p value. An adequate model is nonsignificant; therefore, a p value greater than .01 was used here as criterion. The incremental fit (or goodness of fit) concerns the degree to which the model in question is superior to a baseline model, usually one that specifies no covariances among variables. Larger values indicate greater improvement of the model over a baseline. Recommended incremental fit indicators are the Tucker–Lewis index (TLI) and the comparative fit index (CFI). An optimal value is 1.0 but is seldom obtained in practice (Cheung & Rensvold, 2002). A cutoff score of .95 for each is recommended by Hu and Bentler, although it has been suggested that these are too high and that a cutoff of .90 is reasonable (Cordon & Finney, 2006). A range from .90 (lower bound) to .95 (optimal bound) is used as a criterion in this study.

Finally, the root-mean-square error of approximation (RMSEA), which allows one to test the lack of fit of the sample data to the model, was also included in the analysis. Values closer to zero are optimal, and a cutoff score of .06 for RMSEA has been recommended by Hu and Bentler (1999). For the purposes of this study, a range from .06 (lower bound) to .08 (optimal bound) is used as criterion for goodness of fit.

Analyses of Invariance: Multiple Groups Approach

The assessment of measurement invariance consists of a series of stepwise analyses to determine the extent to which items have equal meaning across different groups of subjects. This type of analysis enhances the examination of construct validity of the test because it provides evidence about construct-irrelevant invariance (Cordon & Finney, 2006). As French and Finch (2006) noted, new tests must meet current validity standards, including the analysis of invariance, which limits the creation of biased measures. Invariance tests also facilitate better interpretation of true differences underlying latent constructs (e.g., Cordon & Finney, 2006).

There are three main steps of measurement invariance: configural, metric, and scalar invariance tests. Configural invariance analysis tests the stability of the hypothesized factor structure across groups by evaluating whether the same indicators (i.e., items) are associated with the same factors across groups (Rensvold & Cheung, 2001). Metric invariance analysis tests the hypothesis of equality of factor loadings. If the model was found variant, examination of each factor and then the items is recommended (Rensvold & Cheung, 2001). Finally, scalar invariance assumes equivalence of intercepts across groups (Rensvold & Cheung, 2001). In other words, it sets constraints to the model to

examine whether respondents with equivalent values on the latent constructs have equivalent values on the items as well.

For the purposes of the screener, two multiple group analyses were performed. The first analysis included the measurement invariance across sex (female vs. male). The second included two groups, distributed by age: a young group, with children ages 6 to 8 years, and an older group, with children ages 9 to 11 years. We hypothesized that the model would be held invariant across gender groups (i.e., there is not evidence supporting developmental sex-related differences in executive functions). We also hypothesized that the model would be invariant across age (i.e., there may be developmental differences in level but not in the structure of the model of executive function itself). Finally, a CFI difference (ΔCFI) $\leq -.01$ was the criterion used for the measurement invariance as recommended by Cheung and Rensvold (2002) and was the criterion used in other studies (French & Finch, 2006; Kim, Cramond, & Bandalos, 2006).

Results

Phases 1 Through 4: Data and Item-Level Screening

The DeCarlo macro (DeCarlo, 1997) identified five cases with significantly large Mahalanobis distances, which were identified as outliers and were eliminated. The overall effective sample size was 2,165 children. Further, the full data set was assessed for univariate normality. The skewness and kurtosis values for most items were within criterion parameters for univariate normality ($|2.0|$ and $|7.0|$, respectively), with the exception of Item 28, which had a value slightly over the cutoff score for skewness (2.2). It is worth noting that all items were significantly correlated ($p < .01$), but correlations ranged from low to modest. Correlations ranged from a minimum of $-.10$ (between Item 3, *makes decisions easily,* and Item 18, *argues when denied own way*) to a maximum of .76 (between Item 12, *is easily distracted from classroom*, and Item 14, *has a short attention span*).

Phase 5: Initial Reliability and Validity Analyses

Internal consistency coefficients ranged from .805 for Problem Solving to .890 for Attentional Control. Alpha coefficients for the Behavioral Control and Emotional Control factors were .842 and .845, respectively. These four coefficients are considered indicators of moderate to high reliability. The Cronbach's-alpha-if-item-deleted feature was used to identify items that did not significantly contribute to the scale. None of the items demonstrated an increase in alpha coefficient if deleted. A lower coefficient for the full screener (.640) was obtained as expected, and it was interpreted as a potential indicator of the multidimensionality of this executive functions measure.

Confirmatory Factor Analysis

To answer the question about the appropriateness of considering the screener a unidimensional versus multidimensional instrument, a series of CFA models were analyzed (see Figure 2).

The first model (Model 1) corresponds to a one-factor/unidimensional model with 25 indicators. The latent construct was called Executive Function, as it presumably measures executive functions as a whole. The model rejection rate in Mplus was the highest (100%), and the model did not converge. Models with two, three, and four factors converged normally. Model 2 had two latent constructs, one called Problem Solving and a second construct referred to here as Behavioral Self-Regulation. This second factor comprised items that originally corresponded to the factors Attentional Control, Behavioral Control, and Emotional Control. A third model (Model 3) comprised three latent constructs: Problem Solving, Attentional Control, and Behavioral/Emotional Control. These last three models demonstrated increasingly better fit in terms of a lower (or less inflated) chi-square, more degrees of freedom, and higher fit indexes. Chi-square differences were observed but not tested because the models were not nested. CFI differences are included in Table 3. Although a small CFI difference between Model 3 and Model 4 was observed, RMSEA and TLI indexes are higher in Model 4. These results support the hypothesis that the group of items selected for this screener yields information about more than one component of executive functions; therefore, it is possible to reject the null hypothesis for unidimensionality. It also demonstrated that the theoretical model proposed (four factors) is not only viable but also testable.

Model modifications. Once the multidimensionality of the model was confirmed, the next analysis consisted of examination

Table 3
Model Variation Analyses

Model	WLSMV χ^2	*df*	CFI	ΔCFI	TLI	RMSEA
Fit indexes for the unidimensional model versus multidimensional models						
Model 1: 25 items, 1 factor	No convergence					
Model 2: 25 items, 2 factors	4,856.138	84	0.827		0.944	0.162
Model 3: 25 items, 3 factors	2,245.639	106	0.923	−0.096	0.980	0.097
Model 4: 25 items, 4 factors	1,899.635	111	0.935	−0.012	0.984	0.086
Exploratory analysis following modification index recommendations						
Model 4: baseline	1,899.635	111	0.935		0.984	0.086
Model 5: Item 20 loading on all	1,469.593*	111	0.951	−0.016	0.988	0.075
Model 6: Item 20 loading on PS and BC	1,550.552	112	0.948	0.003	0.987	0.077
Model 7: Item 20 eliminated	1,403.811	105	0.952		0.988	0.076

Note. WLSMV χ^2 = weighted least squares with mean and variance adjusted chi-squared; *df* = degree of freedom; CFI = comparative fit index; ΔCFI = comparative fit index differences between models; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation.

of the model modification indexes. This step commonly corresponds to an exploratory analysis because these modifications are data driven, or a posteriori, rather than theoretically driven. It has been observed that theory should prevail, as there is an ongoing tendency to overfactorize latent constructs (Frazier & Youngstrom, 2007). A common practice is to negotiate between the two spheres and make decisions on model changes on the basis of indexes but supported by theory and guided by two principles: parsimony and replicability. Modification indexes for Model 4 (from Figure 2), which we call the *baseline model*, suggest allowing Item 20, *acts without thinking*, to load in all factors. It was originally set to load only on the Behavioral Control factor. Table 3 summarizes the CFIs across models.

As reported in Table 3, Model 4 (baseline) was compared against three more models. Model 5 included factor loadings from each factor to Item 20; Model 6 included factor loadings from the Behavioral Control and Problem Solving factors. Finally, in Model 7, the factor loading of Item 20 was constrained to zero, eliminating the relationship between the indicator (Item 20) and the latent constructs. Figure 3 illustrates these models.

Parameters estimates. Table 4 summarizes the factor loadings for each item. Item 20 has a negative estimate for two possible reasons: (a) it shares variance with more than one construct and needs to compensate for this, or (b) it is negatively worded, whereas all the items on Problem Solving are positively worded. Factor loadings represent the strength of the relationship between the construct and the item in terms of shared variance. Standardized values closer to 1.0 are optimal. In case of cross-loadings (e.g., Item 20), a lower estimate was expected. These estimates ranged from .618 to .936, which are significant and explain the overall goodness of fit of the model.

Figure 4 represents the final model configuration and includes the correlations among factors. As can be observed in the figure, the correlations range from $r = -.502$ to $r = -.875$, which are considered moderate to high. Significant moderate-to-high correlations were expected between this screener's four factors because there are strong relationships between the underlying latent construct they represent. It is possible that, as executive functions, these constructs are measuring components of the same underlying (and unifying) latent construct, which could be a second order construct in this model. Negative correlations were observed here and in the interitem matrix. This negativity is explained by the wording orientation of the items. Problem Solving items included positively worded items (with the exception of Item 20, which is negatively worded), and the other three factors used only negatively worded items.

Analyses of Measurement Invariance

Measurement invariance across gender. Results are summarized in Table 5. The CFI difference (ΔCFI) between the configural invariance model and the metric invariance model is $\Delta\text{CFI} = -.005$, which is within the cutoff score limit of $\Delta\text{CFI} \leq -.01$ recommended by Cheung and Rensvold (2002). Moreover, scalar invariance was supported in the current analysis. The difference between the metric and the scalar invariance models is $\Delta\text{CFI} = -.011$, which is reasonably within the cutoff score of $-.01$ (Cheung & Rensvold, 2002). Other fit indexes, such as TLI and RMSEA, also demonstrated adequate goodness of fit.

Measurement invariance across developmental ages. Measurement invariance was not expected across gender groups as much as it would be expected across ages (Becker, Isaac, & Hynd,

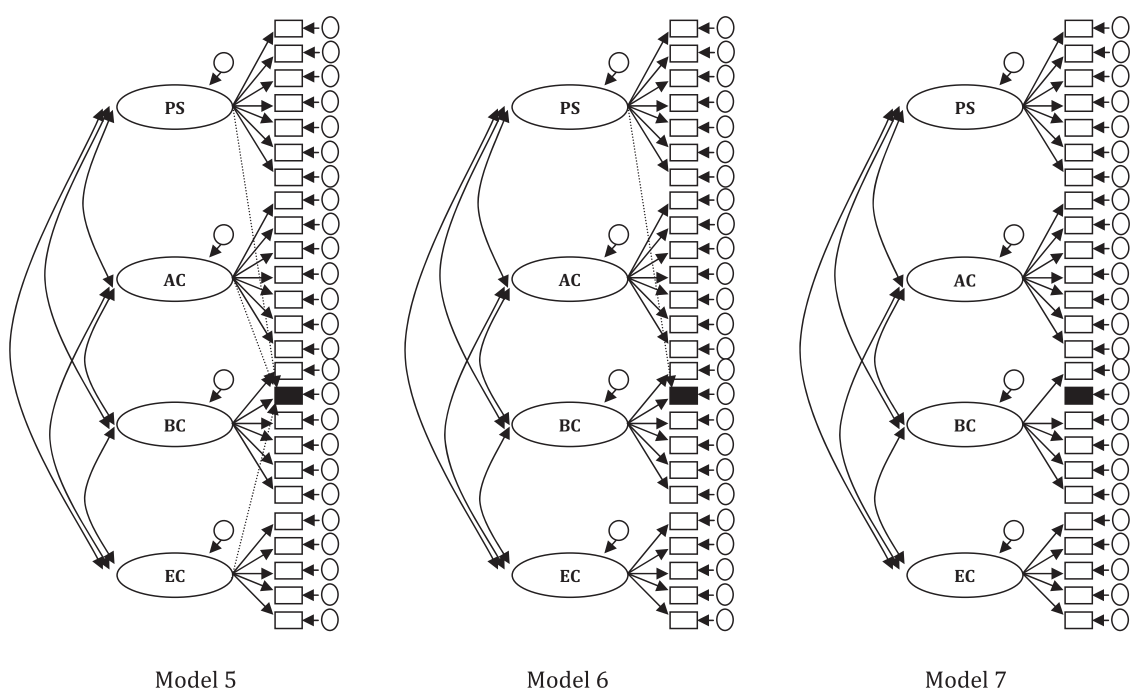

Figure 3. Test for model modifications. PS = Problem Solving; AC = Attentional Control; BC = Behavioral Control; EC = Emotional Control.

Table 4
Factor Loadings

Item	Unstandardized factor loading	Standardized factor loading	*SE*	R^2
Problem Solving				
Item 1	1.000	.740	.000	.547
Item 2	0.977	.723	.023	.523
Item 3	0.901	.666	.022	.444
Item 4	1.164	.861	.022	.742
Item 7	0.997	.738	.024	.544
Item 9	1.265	.936	.023	.877
Item 10	1.207	.893	.023	.797
Item 20	−0.445	−.329	.027	.717
Attentional Control				
Item 11	1.000	.671	.000	.450
Item 12	1.357	.911	.031	.829
Item 13	1.083	.726	.028	.528
Item 14	1.364	.915	.033	.838
Item 15	1.327	.890	.031	.793
Item 16	1.015	.681	.029	.464
Item 17	1.169	.784	.029	.615
Behavioral Control				
Item 19	1.000	.823	.000	.677
Item 20	0.751	.618	.025	.717
Item 21	1.042	.857	.022	.734
Item 22	0.983	.809	.023	.654
Item 23	0.997	.820	.022	.673
Item 24	0.860	.707	.024	.500
Emotional Control				
Item 18	1.000	.884	.000	.782
Item 25	0.842	.744	.018	.554
Item 26	0.871	.770	.018	.593
Item 27	0.853	.755	.018	.569
Item 28	1.011	.893	.017	.798

1987; Davidson, Amso, Anderson, & Diamond, 2006). Configural invariance for the two age groups was met in this study. As reported in Table 5, goodness-of-fit indexes were within the cutoff scores. Specifically, the CFI (.946) was between the lower bound and the optimal bound recommended by Hu and Bentler (1999), the TLI was closer to the optimal level, and the RMSEA was at the cutoff point. Metric invariance was also met in this study. The WLSMV adjusted chi-square was lower than in the adjacent model. Yet, and as expected in CFA models with large sample sizes, the change in chi-square test was significant ($p = .0018$), the CFI was .950, and the ΔCFI was $-.004$ with both CFI values within the cutoff scores. TLI and RMSEA values improved as well. Equality of factor loadings also demonstrated that the strength between the items and the latent constructs was similar across these two age groups. Finally, Scalar invariance was also met by the screener's model. These results are summarized in Table 5.

Discussion

Latent Construct Analysis and Operationalization

In the present study, we used a set of selected items from the BASC-TRS-C (Reynolds & Kamphaus, 1992) to develop a theoretical and psychometrical derivation of a screener for the assessment of executive functions. Four out of the five factors included in a theory-based model recently developed by Garcia-Barrera (2010) were used to scaffold the instrument. These four factors represented four executive functions: A first factor, Problem Solving, included behaviors related to planning and initiation of goal/task attack; three more factors, Attentional Control, Behavioral Control, and Emotional Control, were related to the self-regulatory mechanisms involved in the execution of the programs needed for goal attainment. These four components were treated as latent constructs in the CFA model, and their observed variables were 25 items from the BASC-TRS-C that were believed to serve as indicators of the executive behaviors.

Results of the content validity analysis indicated a few problems with the operationalization of the some of constructs that were addressed. For instance, the Problem Solving factor received an average of an 81% agreement rate (out of 10 items). Although this could mean that some of the items loading in this factor measure construct-irrelevant content, factor loadings for this construct were significant and demonstrated that its indicators explained a large amount of the variance on the latent construct. However, three out of the original 10 items on this factor were eliminated to more accurately represent the target latent construct. These adjustments did not increase internal consistency coefficients; yet, they did increase construct validity as it was observed on model goodness of fit.

It could be also argued that this first factor is attempting to measure too many components of executive function. Because of its intrinsic condition as a screener imbedded in a larger behavior assessment rating scale, it was important to target a diversity of behaviors believed to be executive in nature. We believe that during behavioral regulation mediated by executive systems, some of the most important parts of the process are planning the goal execution, organizing information (input, output), making decisions about the goal/task approach, and programming the action (Zelazo et al., 1997). In this context, the factor Problem Solving could also be recognized as preparation for an action step toward goal execution. Furthermore, in the BRIEF manual, Gioia et al. (2000) reported a similar situation regarding the Plan/Organize scale. They asserted that an empirical analysis of the BRIEF structure demonstrated that these two areas should be integrated into one factor, even though theoretically they could be separated into independent executive functions.

Another caution was raised by the content validity analysis and was related to difficulty in differentiating the placement of items between the Behavioral Control and the Emotional Control factors. Moreover, it could be argued that it is difficult to differentiate these two factors from each other, and this is demonstrated by the small CFI difference (ΔCFI $= -.012$) observed between the three-factor model (Model 3, see Figure 2) and the four-factor model (Model 4). There are several avenues to support the differentiation among some of the factors. The first is related to the conceptualization of two constructs—behavioral inhibition and behavioral self-control—and their theoretical relationship with the construct of emotional self-regulation. Behavioral inhibition refers to the ability to inhibit initial prepotent responses to external cues or events, stop an ongoing response producing a delay period, and protecting this delay from interferences (Barkley, 1997). In contrast, during emotional self-regulation, the delayed prepotent re-

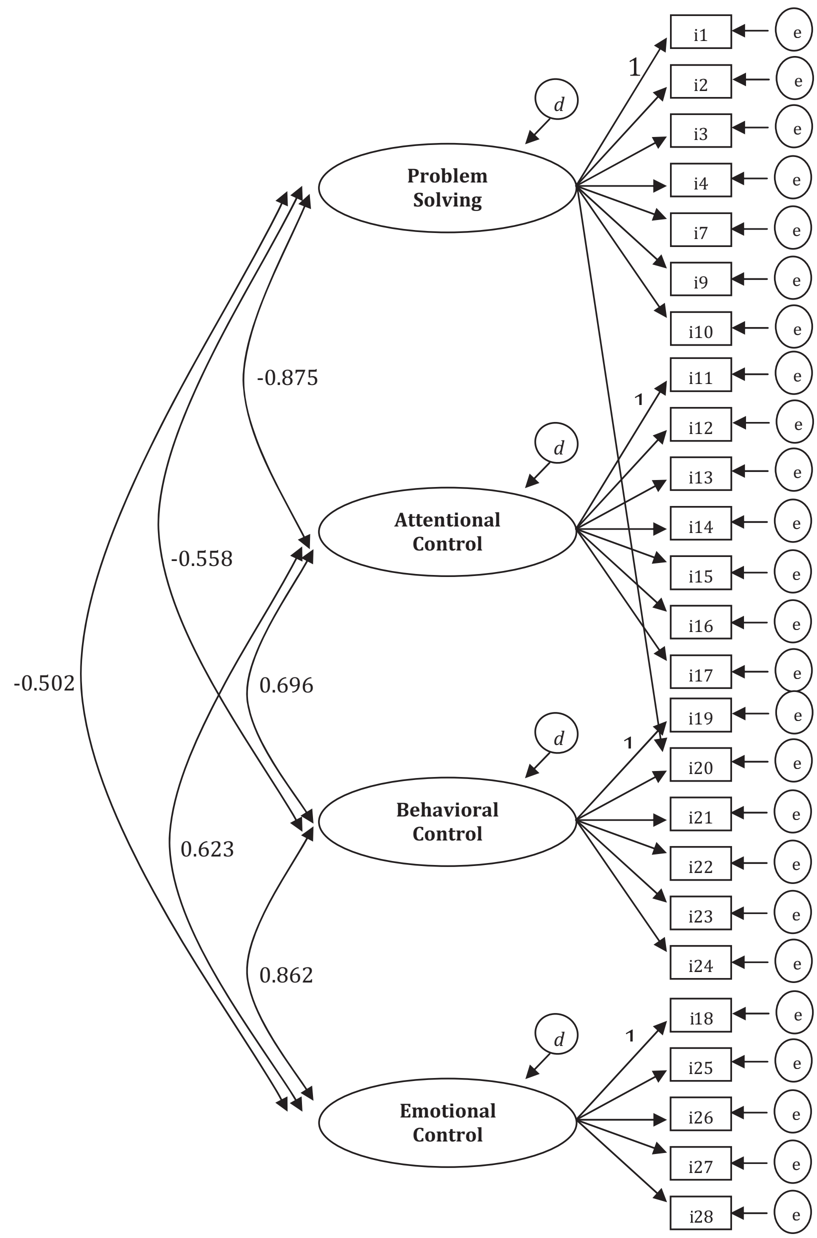

Figure 4. Final confirmatory factor analysis model for the Behavior Assessment System for Children Executive Functions screener. i = item.

sponse corresponds to the expression of emotional reactions elicited by the event, although they may be intrinsically related (Barkley, 1997).

Second, there is research supporting two differentiated brain pathways or circuits from the prefrontal cortex and in relation to self-regulation of behavior and emotion. Rolls (2002) asserted that this regulatory function may be mediated by the involvement of orbitofrontal cortex in the decoding and representation of primary reinforcers, which in turn regulate reward-related and punishment-related behavior. Furthermore, Fuster (2008) affirmed that the medial orbitofrontal cortex has a close connection to limbic structures (i.e., amygdala), through the paralimbic cortex. In this way, although there are common structures and networks involved in the self-regulation of behavior and emotion, it could be reasonable to separate them for measurement purposes.

Third, the debate about the best way to estimate executive functions has shed some light on alternative approaches that add ecological validity to the assessment of executive functions. In this regard, Gioia et al.'s (2000) BRIEF includes two separate scales, the Inhibit scale and the Emotional Control scale, which are integrated on a composite index score denominated Behavioral Regulation Index. It could be argued that the two factors Emotional Control and Behavioral Control should also be integrated

Table 5
Analysis of Measurement Invariance

Model	WLSMV χ^2	$\Delta\chi^2$ (p)	*df*	CFI	ΔCFI	TLI	RMSEA
Measurement invariance for male and female groups							
Step 1: Configural invariance	1,588.278		218	0.947		0.986	0.076
Step 2: Metric invariance	1,445.774	29.506 (.0301)	215	0.952	−0.005	0.988	0.073
Step 3: Scalar invariance	1,119.477	82.052 (.0000)	178	0.963	−0.011	0.988	0.070
Measurement invariance for two age groups							
Step 1: Configural invariance	1,742.818		218	0.946		0.987	0.080
Step 2: Metric invariance	1,635.108	40.425 (.0018)	220	0.950	−0.004	0.988	0.077
Step 3: Scalar invariance	1,255.042	93.759 (.0000)	181	0.962	−0.012	0.989	0.074

Note. WLSMV χ^2 = weighted least squares with mean and variance adjusted chi-square; $\Delta\chi^2$ = chi-square difference test; *df* = degree of freedom; CFI = comparative fit index; ΔCFI = comparative fit index differences between models; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation.

into one factor for the screener purposes. From a measurement perspective, there are some arguments in favor of and against this idea. On the one hand, and in favor of the separation, there is little available research on the BRIEF's Emotional Control and Inhibition scales and in relation to diagnosis of disorders characterized by executive dysfunction (e.g., Jarrat et al., 2005). Most of the research has looked at the BRIEF index scores (e.g., Mahone et al., 2002). In one study, Gioia, Isquith, Retzlaff, and Espy (2002) tested four different CFA models, using maximum likelihood estimation. In this study, a clinical sample of children with mixed diagnoses was assessed with the BRIEF. Gioia et al. reported that for this clinical sample, a three-factor model obtained the best fit indexes. The BRIEF Emotional Control scale loaded in an independent factor from Inhibit and Self-Monitoring. On the other hand, and against the separation, it could be argued that the high and significant correlation between the two factors ($r = -.862$), the small CFI difference between the three-factor model and the four-factor model (ΔCFI = .012), and the lower agreement rates by the panel make it necessary to consider further analysis of the current factorization. In this regard, Frazier and Youngstrom (2007) recalled the importance of parsimony when developing tests, in terms of the number of factors they attempt to address. They also asserted that "recent commercial tests of cognitive ability are not adequately measuring the number of factors they are purported to measure by test developers" (Frazier & Youngstrom, 2007, p. 180).Finally, they reminded test developers of the recent increase of overfactorization. It would be advisable to further examine a four-factor versus a three-factor model across different samples.

One more factor should be addressed: Attentional Control. Starting at an item level, interrater agreement rates for the items in this factor were higher, and a reasonable consensus was achieved. We noticed during the item screening analysis that some of the items within this factor had the highest correlations across the matrix. In other words, it could be argued that some of the items were redundant in terms of behavioral indicators on a screening instrument. A look into the statistical analysis for this construct may serve as a counterargument for this assumption. Factor loadings were generally significant, ranging from .671 to .915, and were stable across gender and developmental ages. The R^2 values, which indicate the contributions of each item to the shared latent factor variance, ranged from .450 to .838 and are within the significant range. Moreover, alpha coefficients did not significantly increase when any of the seven items were deleted.

Furthermore, and at the construct level, it could be argued that Attentional Control should be considered a second-order or higher level self-regulatory ability, responsible for the modulation of behavior. This assumption comes from the works of Posner and Rothbart (1998) and Rueda et al. (2005), who also labeled this construct *executive attention* or the *executive control system.* These authors have called this system the *attentional organ*. In its complexity, this organ appears to be a reasonable theoretical framework for the study of executive functions, if one assumes that this organ regulates executive functions or that executive functions are not more than attentional functions in nature. The Attentional Control factor includes several items that are related to the functions of Posner and Rothbart's and Rueda et al.'s attentional mechanism. However, as it is directly presented by these authors, it would be difficult to use behavioral indicators on a screening instrument (e.g., ratings) to operationalize it as a latent construct, whereas it has been successfully assessed with laboratory and performance-based measures. Therefore, the aim of this factor is to estimate the executive ability to regulate attentional functions, such as focusing, sustaining, and shifting.

Model Modifications and Item Analysis

At an item level, one more issue regarding the factorial structure of the screener is worth discussing here. Modification indexes produced by Mplus recommended the inclusion of a factor loading from all the constructs to Item 20, *Acts without thinking*. This item was originally conceived as a Behavioral Control item, and the panel of experts agreed 100% about its contribution to the operationalization of this factor. The question here is, What is it about this item that may or may not make it contribute to all constructs? A look to this item reveals that it captures the essence of executive functions, in the sense that a behavior such as acting without thinking not only represents the lack of self-control but also the lack of planning and organization of the action prior to its execution or, in the best scenario, the inefficient preparation for the initiation of action. An excellent example of this type of behavior is observed in children with ADHD (Barkley, 1997). Similar arguments could be made for the inclusion of factor loadings from the Attentional Control and Emotional Control factors to Item 20.

One could present the argument that attentional systems are regulated by the intentionality and the drive to attain a goal (Fuster, 2008) or that acting without thinking may relate to an automatic-driven (posterior) more than voluntary-driven (anterior) focusing of attentional systems (Rueda et al., 2005) on information that is irrelevant or relevant depending on the context (e.g., survival). However, in the context in which this item is presented, acting without thinking suggests a poor, problematic, and inefficient behavior related to overall lack of control.

Unidimensionality Versus Multidimensionality

One more issue worth discussing here refers to single-factor model versus multiple-factor model comparisons. Results clearly pointed out the need for expansion on the factorization, from a nonconverging model (unidimensional) to three- and four-factor models that not only converged but also demonstrated adequate goodness of fit. Earlier, we discussed the appropriateness of using a four- versus a three-factor model. Yet, the question is how do these results impact the current ways of assessing executive functions?

This issue can be addressed from different angles. From a statistical standpoint, it appeared appropriate to factorize the set of 25 items into three or four components. We based the decision to utilize a four-factor model on theory, both a priori—as the four constructs were originally defined before examining the fit of the model—and a posteriori—as some statistically driven (or data-driven) modifications were applied to the model. From a theoretical standpoint, the construct validity analyses demonstrated the pertinence of including four components of executive functions in the screener. As we stated earlier, it is difficult to confine the definition of *executive function* into one statement (Elliot, 2003), as much as it is difficult to avoid numerating its components while defining the construct (Miyake et al., 2000) and to list all of the possible components under the executive functions umbrella (Baddeley, 1996). We decided to utilize the term *executive functions* as a manifest of an understanding of this latent construct in terms of its diversity, believing that (as is the case for "g") executive function could be approached as a second-order latent construct that unifies the diversity of its components. The invariant sustainability of the model across different groups may serve as evidence in favor of this statement. In other words, it seems appropriate to approach the assessment of executive function from its components, through indicators, and from there to estimate the latent construct as a composite.

Having said that, the following question may come to mind: Is the composite more useful than the analysis of its components? This is a difficult question to address, and it has been largely discussed in the intelligence testing literature. It is known from analysis of overall composite scores versus part scores that the full composite is more stable (Canivez & Watkins, 2001; Neisser et al., 1996), has the best predictive validity (Hunter & Hunter, 1984), and has a large amount of scientific and theoretical support. In contrast, we also recognize that part scores (e.g., indexes, scales, factors) allow clinicians to perform profile analysis, in which patterns of subtest scores assessing individuals' strengths and weaknesses are interpreted. Part scores facilitate diagnosis and the selection of appropriate interventions (Glutting, McDermott, Konold, Snelbaker, & Watkins, 1998).

According to Kamphaus (2001), profile analyses can be of two types: normative and ipsative. Normative subtest profiles compare an individual's subtest scores with those of a norm-referenced group, whereas ipsative analyses interpret intraindividual differences. Despite their popularity, ipsative profile analysis is not supported by empirical evidence. However, it would be recommended in clinical practice and when working with special populations (e.g., bilingual assessment) and for screening purposes. Moreover, it appears appropriate to create a composite score for this screener, which would be multidimensional but would also underline the unity of the construct.

Measurement Invariance and Construct Validity

We performed two independent multiple-group CFA analyses in this study to identify measurement equivalence across groups. The first analysis included a stepwise examination of configural, metric, and scalar invariance across female and male subsamples. The second analysis included the same three examinations across two age groups: children 6 to 8 years old (young group) and children 9 to 11 years old (older group). Configural invariance was supported by the model in both cases, which demonstrates that the same four-factor structure underlying the screener is being measured across groups, in that "the same indicators are associated with the same factors for both groups" (Resvold & Cheung, 2001, p. 29). Metric invariance was also met. According to Rusticus and Hubley (2006), if metric invariance is met, "the measure may be used to examine structural relationships or correlations between the construct of interest and other constructs across groups" (p. 828). Because of the overall fit of the model, no further metric invariance testing of each factor or of the items was necessary. Finally, the model also supported the strongest type of examination, scalar invariance. However, a careful examination of the modification indexes produced after the scalar invariance test suggested some meaningful model variations at the item level and across groups. For instance, allowing a factor loading from Item 10, *Is well organized*, to all factors could have some impact on overall fit in the female group but not in the male group. Similarly, allowing the same item to load on the Behavioral Control and Emotional Control factors would have an impact on model fit for the younger group but not for the older group. These findings are of great importance for future research involving this screening instrument.

Furthermore, and although gender differences were not expected in this model, it could be argued that differences across developmental ages should be identified by the model, because of results obtained in experimental research in the development of executive functions (e.g., Becker et al., 1987; Davidson et al., 2006; Diamond, 2002; Zelazo et al., 2003). These differences could be identified in two ways: shape and level. Configural, metric, and scalar invariance examinations serve as tests for evaluating differences in shape (factorial and construct structure). As we reported, measurement invariance was supported. Differences in level indicate whether one group has a greater amount of latent construct than the other; in this case, it would be necessary to perform a latent means difference test to identify level differences.

Clinical Implications, Limitations, and Future Research

This study demonstrated a psychometrically and theoretically sound screening tool for executive functioning derived from a large pool of items, such as that provided by the BASC system. Because this instrument was designed within the larger BASC-TRS-C, its implementation should not be difficult; scores on the four scales can be derived following a BASC administration. Although screeners should not be used as diagnostic tools, brief screeners are optimal and relatively inexpensive tools for early identification of executive difficulties. Yet, there are some limitations that should be addressed before recommending this instrument for clinical or research use. First, analysis of convergent validity would be necessary to establish whether this instrument is comparable and competitive with other executive behavior screeners, such as the BRIEF. Moreover, and because of the early identification of low correlations between the BRIEF Parent and Teacher scales with performance-based tests of executive functions (e.g., Anderson, Anderson, Northam, Jacobs, & Mikiewicz, 2002; Vriezen, & Pigott, 2002), it is extremely important to evaluate the relationships between the executive behaviors evaluated by this screener and those evaluated by performance-based tests. Second, it is necessary to create *T* scores for each one of the four constructs imbedded in the model, so that norm comparisons are possible. Third, introduction of the composite score would be optimal for clinical applications and cross-instruments comparisons. Fourth, further analysis of the clinical validity of the instrument would enhance its utility as an assessment instrument. We recommend that a large clinical database be used to examine the instrument's sensitivity and sensibility and, therefore, its discriminant accuracy rate. Furthermore, it would be relevant to derive a screener similar to that from the BASC Parent Rating scale. Fifth, because of to its nature, this screener was conceived within the limited frame of available items from the BASC. Therefore, there are important executive functions (e.g., updating working memory; Garcia-Barrera, 2010; Miyake et al., 2000) that could not be included in the model and are of great importance to the assessment of this construct. Updating of working memory, in particular, is a difficult executive component to capture on a behavior scale. Items associated with memory in general may add impurity to the construct. We recommend the use of clinical assessment tools for working memory to obtain reliable information regarding updating working memory representations (e.g., *N*-back tests).

In summary, clinical validity analysis of the interpretations derived from the screener scores are imperative, and future research should include convergent and concurrent validity analyses with comparisons to the BRIEF and to performance-based gold standard measures of executive functions as well as cross-cultural validation of the model.

References

Anderson, V. (2001). Assessing executive functions in children: Biological, psychological, and developmental considerations. *Pediatric Rehabilitation, 4,* 119–136.

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships between cognitive and behavioral measures of executive function in children with brain disease. *Child Neuropsychology, 8,* 231–240. doi:10.1076/chin.8.4.231.13509

Ardila, A. (2008). On the evolutionary origins of executive functions. *Brain and Cognition, 68,* 92–99. doi:10.1016/j.bandc.2008.03.003

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49*A, 5–28.

Barkley, R. A. (1997). *ADHD and the nature of self-control.* New York, NY: Guilford Press.

Barringer, M. S., & Reynolds, C. R. (1995, November). *Behavior ratings of frontal lobe dysfunction.* Paper presented at the annual meeting of the National Academy of Neuropsychology, San Francisco.

Becker, M. G., Isaac, W., & Hynd, G. W. (1987). Neuropsychological development of nonverbal behaviors attributed to "frontal lobe" functioning. *Developmental Neuropsychology, 3,* 275–298. doi:10.1080/87565648709540381

Borkowski, J. G., & Burke, J. E. (1996). Theories, models, and measurements of executive functioning: An information processing perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 235–261). Baltimore, MD: Brookes.

Canivez, G. L., & Watkins, M. W. (2001). Long-term stability of the Wechsler Intelligence Scale for Children—Third Edition among students with disabilities. *School Psychology Review, 30,* 438–453.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28,* 595–616. doi:10.1207/s15326942dn2802_3

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

Clark, C., Prior, M., & Kinsella, G. (2002). The relationship between executive function abilities, adaptive behaviour, and academic achievement in children with externalising behaviour problems. *Journal of Child Psychology and Psychiatry, 43,* 785–796. doi:10.1111/1469-7610.00084

Conners, C. K. (1997). *Conners' Rating Scales–Revised: Technical manual.* North Tonawanda, NY: Multi-Health Systems.

Cordon, S. L., & Finney, S. J. (2006). *Measurement invariance of the Mindful Attention Awareness Scale across adult attachment style.* Unpublished manuscript.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York, NY: Holt, Rinehart & Winston.

Damasio, A. R. (1995). On some functions of the human prefrontal cortex. *Annals of the New York Academy of Sciences, 769,* 241–251. doi:10.1111/j.1749-6632.1995.tb38142.x

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44,* 2037–2078. doi:10.1016/j.neuropsychologia.2006.02.006

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2,* 292–307. doi:10.1037/1082-989X.2.3.292

Denckla, M. B. (2007). Executive function: Binding together the definitions of attention-deficit/hyperactivity disorder and learning disabilities. In L. Melltzer (Ed.), *Executive function in education* (pp. 5–18). New York, NY: Guilford.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195134971.003.0029

Elliott, R. (2003). Executive functions and their disorders. *British Medical Bulletin, 65,* 49–59. doi:10.1093/bmb/65.1.49

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,* 430–457. doi:10.1207/S15328007SEM0803_5

Espy, K. A. (2004). Using developmental, cognitive, and neuroscience ap-

proaches to understand executive control in young children. *Developmental Neuropsychology, 26,* 379–384. doi:10.1207/s15326942dn2601_1

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

Finney, S. J., & DiStefano, C. (2006). Nonnormal and categorical data in structural equation models. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling. A second course* (pp. 269–314). Greenwich, CT: Information Age.

Flanagan, K. S., Bierman, K. L., & Kam, C. M. (2003). Identifying at-risk children at school entry: The usefulness of multibehavioral problem profiles. *Journal of Clinical Child and Adolescent Psychology, 32,* 396–407. doi:10.1207/S15374424JCCP3203_08

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9,* 466–491. doi:10.1037/1082-989X.9.4.466

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence, 35,* 169–182. doi:10.1016/j.intell.2006.07.002

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling, 13,* 378–402. doi:10.1207/s15328007sem1303_3

Fuster, J. M. (2002). Frontal lobe and cognitive development. *Journal of Neurocytology, 31,* 373–385. doi:10.1023/A:1024190429920

Fuster, J. M. (2008). *The prefrontal cortex* (4th ed.). London, United Kingdom: Academic Press.

Garcia-Barrera, M. A. (2010). *Assessing executive functions in the schools.* Manuscript in preparation.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Functions.* Odessa, FL: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Retzlaff, P. D., & Espy, K. A. (2002). Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function (BRIEF) in a clinical sample. *Child Neuropsychology, 8,* 249–257. doi:10.1076/chin.8.4.249.13513

Glutting, J. J., McDermott, P. A., Konold, T. R., Snelbaker, A. J., & Watkins, M. W. (1998). More ups and downs of subtest analysis: Criterion validity of the DAS with an unselected cohort. *School Psychology Review, 27,* 599–612.

Grace, J., & Malloy, P. (2001). *Frontal Systems Behavior Scale (FrSBe): Professional manual.* Lutz, FL: Psychological Assessment Resources.

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.

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. (2002). Executive functions and development: Emerging themes. *Infant and Child Development, 11,* 201–209. doi:10.1002/icd.297

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

Hunter, J. E., & Hunter, R. V. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin, 96,* 72–98. doi:10.1037/0033-2909.96.1.72

Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology, 26,* 403–422. doi:10.1207/s15326942dn2601_3

Jarratt, K. P., Riccio, C. A., & Siekierski, B. M. (2005). Assessment of attention deficit hyperactivity disorder (ADHD) using the BASC and BRIEF. *Applied Neuropsychology, 12,* 83–93. doi:10.1207/s15324826an1202_4

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

Kamphaus, R. W. (2001). *Clinical assessment of child and adolescent intelligence* (2nd ed.). Boston, MA: Allyn & Bacon.

Kamphaus, R. W., Thorpe, J. S., Winsor, A. P., Kroncke, A. P., Dowdy, E. T., & VanDeventer, M. (2007). Development and predictive validity of a teacher screener for child behavioral and emotional problems at school. *Educational and Psychological Measurement, 67,* 342–356. doi:10.1177/00131644070670021001

Kim, K. H., Cramond, B., & Bandalos, D. L. (2006). The latent structure and measurement invariance of scores on the Torrance Test of Creative Thinking–Figural. *Educational and Psychological Measurement, 66,* 459–477. doi:10.1177/0013164405282456

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Lehto, J. E. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology, 49*A, 29–50.

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

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

Lochman, J. E. (1995). Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology, 63,* 549–559. doi:10.1037/0022-006X.63.4.549

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology.* New York, NY: Basic Books.

Mahone, E. M., Zabel, T. A., Levey, E., Verda, M., & Kinsman, S. (2002). Parent and self-report ratings of executive function in adolescents with myelomeningocele and hydrocephalus. *Child Neuropsychology, 8,* 258–270. doi:10.1076/chin.8.4.258.13510

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury, 18,* 1067–1081. doi:10.1080/02699050410001672387

Mirsky, A. F., Anthony, B. J., Duncan, C. C., Ahearn, M. B., & Kellam, S. G. (1991). Analysis of the elements of attention: A neuropsychological approach. *Neuropsychology Review, 2,* 109–145. doi:10.1007/BF01109051

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,* 169–183. doi:10.1055/s-2000-7563

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,* 621–640. doi:10.1037/0096-3445.130.4.621

Muthén, B. O., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes, 22*. Retrieved from http://www.statmodel.com/examples/webnote.shtml

Muthén, L. K., & Muthén, B. O. (2006). *Mplus user's guide.*(4th ed.). Los Angeles, CA: Authors.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Ceci, S. J., & Al, E. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101. doi:10.1037/0003-066X.51.2.77

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

Pineda, D. A., Kamphaus, R. W., Mora, O., Restrepo, M. A., Puerta, I. C., Palacio, L. G., . . . Holguin, J. A. (1999). Sistema de evaluación multi-

dimensional de la conducta. Escala para padres de niños de 6 a 11 años, versión colombiana. [A system of multidimensional behavior assessment. A scale for parents of children from 6 to 11 years old Colombian version]. *Revista de Neurologia, 28,* 672–681.

Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philosophical Transactions of the Royal Society of London: Series B. Biological Sciences, 353,* 1915–1927. doi:10.1098/rstb.1998.0344

Rensvold, R. B., & Cheung, G. (2001). Testing for metric invariance using structural equation models: Solving the standardization problem. In C. A. Schriesheim & L. L. Neider (Eds.), *Equivalence in measurement* (Vol. 1, pp. 25–50). Greenwich, CT: Information Age.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children (BASC*). Circle Pines, MN: American Guidance Services.

Reynolds, C. R., & Kamphaus, R. W. (2002). *The clinician's guide to the Behavior Assessment System for Children.* New York, NY: Guilford Press.

Riccio, C. A., Hall, A. J., Morgan, A., Hynd, G. W., Gonzalez, J. J., & Marshall, R. M. (1994). Executive function and Wisconsin Card Sorting Test: Relationship with behavioral rating and cognitive ability. *Developmental Neuropsychology, 10,* 215–229. doi:10.1080/87565649409540580

Rolls, E. T. (2002). The functions of the orbitofrontal cortex. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 354–375). New York, NY: Oxford University Press.

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. *Journal of Neuropsychiatry and Clinical Neurosciences, 14,* 377–405. doi:10.1176/appi.neuropsych.14.4.377

Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology, 28,* 573–594. doi:10.1207/s15326942dn2802_2

Rusticus, S. A., & Hubley, A. M. (2006). Measurement invariance of the Multidimensional Body-Self Relations Questionnaire: Can we compare across age and gender? *Sex Roles, 55,* 827–842. doi:10.1007/s11199-006-9135-7

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

Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society of London: Series B. Biological Sciences, 351,* 1405–1411.

Sullivan, J. R., & Riccio, C. A. (2006). An empirical analysis of the BASC Frontal Lobe/Executive Control scale with a clinical sample. *Archives of Clinical Neuropsychology, 21,* 495–501. doi:10.1016/j.acn.2006.05.008

Tsujimoto, S. (2008). The prefrontal cortex: Functional and neural development during early childhood. *The Neuroscientist, 14,* 345–358. doi:10.1177/1073858408316002

Vriezen, E. R., & Pigott, S. E. (2002). The relationship between parental report on the BRIEF and performance-based measures of executive function in children with moderate to severe traumatic brain injury. *Child Neuropsychology, 8,* 296–303. doi:10.1076/chin.8.4.296.13505

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (1998). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation, 8,* 213–228. doi:10.1080/713755570

Zelazo, P. D., Carter, A., Reznick, J. S., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1,* 198–226. doi:10.1037/1089-2680.1.2.198

Zelazo, P., Müeller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function. In W. F. Overton (Series Ed.), The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68,* 1–27.

Received April 4, 2009
Revision received March 2, 2010
Accepted July 1, 2010 ■