3-1-2018

Sensitivity and Specificity of an Executive Function Screener at Identifying Children With ADHD and Reading Disability.

Justin E Karr

Michelle Y Kibby

Audreyana C Jagger-Rickels

Mauricio A Garcia-Barrera

Sensitivity and Specificity of an Executive Function Screener at Identifying Children with ADHD and Reading Disability

Justin E. Karr

University of Victoria

Michelle Y. Kibby & Audreyana C. Jagger-Rickels

Southern Illinois University – Carbondale

Mauricio A. Garcia-Barrera

University of Victoria

Author Note

Justin E. Karr, Department of Psychology, University of Victoria; Michelle Y. Kibby, Department of Psychology, Southern Illinois University – Carbondale; Audreyana C. Jagger-Rickels, Department of Psychology, Southern Illinois University – Carbondale; and Mauricio A. Garcia-Barrera, Department of Psychology, University of Victoria. Justin E. Karr is a Vanier Canada Graduate Scholar and thanks the Natural Sciences and Engineering Research Council of Canada (NSERC) for their support of his graduate studies. Data collection for this study was funded in part by grants awarded to Dr. Kibby from NIH/NICHD (R03 HD048752, R15 HD065627). The views presented here do not necessarily reflect the opinions of NIH/NICHD. Correspondence concerning this article via post should be addressed to Mauricio A. Garcia-Barrera, Department of Psychology, University of Victoria, P.O. Box 1700 STN CSC, Victoria, British Columbia, Canada V8W 2Y2. Electronic correspondence should be addressed to Justin E. Karr (jkarr@uvic.ca).

Abstract

Objective: This study evaluated the sensitivity/specificity of a global sum score (GSS) from the Behavior Assessment System for Children, Second Edition, Executive Function screener (BASC-2-EF) at classifying children with/without ADHD and/or reading disability (RD).

Method: The BASC-2 Teacher/Parent Rating Scales (TRS/PRS) were completed for children (8-12 years-old; 43.1% female) with no diagnosis (n=53), RD (n=34), ADHD (n=85), co-morbid RD/ADHD (n=36), and other diagnoses (n=15). Receiver Operating Characteristic (ROC) Curve analyses evaluated the sensitivity/specificity of the BASC-2-EF GSS at discriminating between children with/without ADHD or RD.

Results: Area Under the Curve (AUC) scores indicated the sensitivity/specificity of the BASC-2-EF GSS at discriminating between children with/without ADHD (TRS: AUC=.831, p<.001; PRS: AUC=.919, p<.001), with/without RD (TRS: AUC=.724, p=.001; PRS: AUC=.615, p=.101), and with ADHD or RD through post-hoc analysis (TRS: AUC=.674, p=.006; PRS: AUC=.819, p<.001).

Conclusion: The findings support utilizing the BASC-2-EF GSS when differentiating ADHD from RD and typical development.

A rich body of research has evaluated the utility of the Behavior Assessment System for Children (BASC) at assessing executive functions (Duggan, Garcia-Barrera, & Müller, 2016; Garcia-Barrera, Duggan, Karr, & Reynolds, 2014; Garcia-Barrera, Kamphaus, & Bandalos, 2011; Garcia-Barrera, Karr, Duran, Direnfeld, & Pineda, 2015; Garcia-Barrera, Karr, & Kamphaus, 2013; Sadeh, Burns, & Sullivan, 2012). Using the BASC-2 Teacher and Parent Rating Scales for Children (TRS-C and PRS-C, respectively), a past derivation study established the reliability and validity of the BASC-2-EF, a four-factor screener embedded in each form that effectively measured four executive-related constructs: problem solving, attentional control, behavioral control, and emotional control (Karr & Garcia-Barrera, 2017). The four-factor measurement model of the BASC-2-EF showed optimal statistical fit, and each factor had excellent internal consistency.

The Behavioral and Emotional Screening System from the BASC-2 was one of the most frequently administered questionnaires in neuropsychological practice (Rabin, Paolillo, & Barr, 2016), and its popularity in research settings was also quite evident, with numerous studies using scales from the BASC-2 as outcome measures in research designs (e.g., Graziano, McNamara, Geffken, & Reid, 2013; Mahan & Matson, 2011). Considering its research popularity, BASC-2 data may be available to researchers through numerous archival datasets, making research on the BASC-2-EF relevant to future research studies involving secondary data analysis (Karr & Garcia-Barrera, 2017), despite the BASC now being in its third edition (BASC-3; Reynolds & Kamphaus, 2015). Aside from executive function measurement more generally, a more specific area where the BASC-2-EF may be useful to researchers is through the assessment of ADHD and evaluation of executive functions among children with ADHD. Although the BASC was not originally developed as a diagnostic measure for ADHD, the instrument has demonstrated clinical utility in aiding its diagnosis (e.g., Jarratt, Riccio, & Siekierski, 2005; Pineda et al., 2005). One previous evaluation of the original BASC found measurement invariance of the BASC-EF screener between children with and without ADHD, and also found ADHD status was a significant predictor of all four factors measured by the BASC-EF (Garcia-Barrera et al., 2015). Thus, the information gathered from the BASC-2 may assist researchers when making diagnostic decisions about participants in secondary data analysis. Considering the relationship between executive dysfunction, ADHD, and functional impairment (Biederman et al., 2004; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005), the limited utility of traditional cognitive performance measures for the diagnosis of ADHD (Pineda, Puerta, Aguirre, Garcia-Barrera & Kamphaus, 2007), and the need to use a multidimensional assessment approach that includes performance-based tests and behavioral ratings (Garcia-Barrera & Kamphaus, 2006), the BASC-2-EF and later iterations (i.e., BASC-3) could assist with such diagnostic decision making, just as other neuropsychological measures of executive functions have demonstrated diagnostic utility (Holmes et al., 2010). Specifically, it has been proposed that ratings of executive behavior may measure a different type of executive function than what performance-based measures examine in ADHD assessment (Toplak, West, & Stanovich, 2013); thus, the BASC-2 EF screener may facilitate the clinical description of four components of executive behavior as parents and teachers observed them, helping characterize the clinical case beyond the diagnosis of ADHD.

The current study consisted of a secondary data analysis of BASC-2 data collected among typically developing children and children with neurodevelopmental diagnoses, specifically ADHD and reading disability. Although the original BASC-EF was validated in a sample with ADHD (Garcia-Barrera et al., 2015), the BASC-2-EF has only been derived using the BASC-2 normative data, consisting entirely of typically developing children (Karr & Garcia-

Barrera, 2017), and it has yet to be clinically validated. Following previous research on the BASC-EF, the current study aimed to (a) replicate the four-factor model of the BASC-2-EF using a mixed clinical sample, (b) identify the sensitivity and specificity of the BASC-2-EF at classifying children with ADHD and RD diagnoses, and lastly (c) produce provisional diagnostic cutoff scores for use in diagnostic decision making.

Method

Participants and Materials

Participants included 223 children (43.1% female; Age: $\bar{x}$=9.49, SD=1.35, range: 8-12 years; Ethnicity: 85.8% Non-Hispanic Caucasian, 4.9% African-American, 2.7% Hispanic, and 6.7% of other ethnicities including multiracial children; Maternal Education: 52.5% with a Bachelor's degree or higher) who took part in larger funded projects examining neuropsychological variables associated with RD and ADHD (Kibby, Lee & Dyer, 2014; Kibby et al., 2015). They consisted of typically developing children (*n*=53) as well as children with RD (*n*=34), ADHD (*n*=85), co-morbid RD/ADHD (*n*=36), and other diagnoses (*n*=15). These groups did not differ significantly based on any demographic variables.

The diagnostic process for ADHD and RD are explained in detail in previous articles (Kibby et al., 2014, 2015). ADHD diagnosis was determined by a clinical neuropsychologist through a three-part process including a parent interview, a diagnostic questionnaire inquiring about the presence of ADHD and other disorders using DSM-IV criteria, and the BASC-2 to assess symptom severity for ADHD and other disorders. Participants without the six symptoms required for diagnosis according to the DSM-IV, but with sufficient symptom severity and impairment, were given an ADHD Not Otherwise Specified diagnosis if they obtained a T-Score ≥60 on the BASC-2 Attention Problems/Hyperactivity subscales. For participants included in the current analysis, the BASC-2-TRS-C and PRS-C were completed by teacher and parent raters, respectively. The sample differs slightly from previous studies, as some of the larger sample had data from the first edition of the BASC and not the BASC-2 and prior studies often focused specifically on RD and/or ADHD and excluded participants with other diagnoses. Both BASC-2 rating scales are designed as broadband measures for the behavioral and emotional evaluation of children ages 6 to 11 years. The BASC-2-TRS-C and PRS-C include 139 and 160 items, respectively, rated on a 4-point Likert-type scale ranging from 1 (*never*) to 4 (*almost always*). The BASC-2-EF included 33 items from the TRS-C and 20 items from the PRS-C (Karr & Garcia-Barrera, 2017). The exact BASC-2 items included in each scale and their factor assignments are presented in Table 1.

Statistical Analyses

The replication of the screener involved a confirmatory factor analysis of a one-factor, two-factor, three-factor and four-factor measurement model conducted in MPlus v7.4 (Muthén & Muthén, 2015), as performed in the original derivation of the BASC-2-EF (Karr & Garcia-Barrera, 2017). The four-factor model was hypothesized to be the best fitting model, but in addition to replicating the four-factor model, a second-order factor model was also calculated, where all first-order factors loaded on a global executive function factor. To estimate the reliability of each factor, Cronbach's α was calculated for each factor using IBM SPSS v.22.

As the BASC-2 items qualify as polytomous, the data was specified as categorical in MPlus. Model fit was evaluated by examining a set of fit indices, including the χ^2 goodness-of-fit test, the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the root-mean-square error of approximation (RMSEA). The χ^2 was not interpreted as the sample size was reasonably large (Tanaka, 1987). In turn, alternative fit indices were considered to assess the fit of the model. A CFI/TLI at or above .90 indicated adequate fit (Bentler & Bonett, 1980), while a CFI/TLI at or above .95 indicated optimal fit (Hu & Bentler, 1999). RMSEA values below .10 indicated good fit, whereas values below .05 indicated excellent fit (Steiger, 1989). Models were compared based on change in CFI, with an increase in CFI at or above .01 indicating significant improvement in model fit (Cheung & Rensvold, 2002).

A global sum score (GSS) of all items included in the BASC-2-EF was calculated for both the TRS-C and PRS-C forms. When calculating the BASC-2-EF GSS for each form, all items responses were coded as *never* (1), *sometimes* (2), *often* (3) and *almost always* (4). A set of items were reverse coded (i.e., 1=4, 2=3, 3=2, 4=1), with these items listed in Table 1. A higher GSS meant greater executive dysfunction. These sum scores were evaluated through a Receiver Operating Characteristic (ROC) Curve analysis to determine the sensitivity and specificity of the screener at discriminating between ADHD or RD and children without any diagnosis. A post-hoc ROC analysis evaluated the sensitivity and specificity of the GSS at distinguishing between children with ADHD and RD. This analysis served to determine if the screener was sufficiently sensitive to differentiate between different neurodevelopmental diagnoses. This resulted in six ROC analyses in total (i.e., ADHD vs. controls, RD vs. controls, and ADHD vs. RD, with each analysis conducted for both the TRS-C and PRS-C). Participants with co-morbid diagnoses were not included in ROC curve analyses to determine the sensitivity and specificity of the screener at differentiating between specific diagnostic categories (i.e., ADHD, RD) and healthy children. The Area Under the Curve (AUC) was interpreted for each analysis, and where the GSS presented with a good or excellent AUC (i.e., good = .80-.90; excellent = .90-1.00), a cutoff score for provisional diagnosis was calculated based on Youden’s J statistic (Youden, 1950).

Results

Table 2 presents the model fit results for all measurement models evaluated. In terms of statistical fit, the four-factor model showed the best fit for both scales (TRS-C: CFI=.942, RMSEA=.087; PRS-C: CFI=.960, RMSEA=.080). The second-order factor model showed a minimal decrement in model fit (TRS-C: CFI=.935, RMSEA=.093; PRS-C: CFI=.948, RMSEA=.091). In the interest of a parsimonious interpretation, the second-order factor model was considered the accepted model, with factor loadings provided for both the TRS-C and PRS-C in Table 1. With these final item assignments, the internal consistency (i.e., Cronbach's α) for all items showed excellent reliability (TRS-C=.952; PRS-C=.913), while each subscale also presented good to excellent internal consistency, including problem solving (TRS-C=.870; PRS-C=.820), attentional control (TRS-C=.927; PRS-C=.902), behavioral control (TRS-C=.932; PRS-C=.789), and emotional control (TRS-C=.759; PRS-C=.682).

The ROC analysis produced an Area Under the Curve (AUC) that corresponded to the sensitivity and specificity of the BASC-2-EF GSS at discriminating between children with and without ADHD (TRS-C: AUC=.831 [95% C.I.: .761, .901], *p*<.001; PRS-C: AUC=.919 [.858, .979], *p* <.001) and children with and without RD (TRS-C: AUC=.724 [.612, .835], *p*=.001; PRS-C: AUC=.615 [.481, .748], *p*=.101). The post-hoc analysis evaluated the sensitivity of the BASC-2-EF GSS at differentiating between ADHD and RD (TRS-C: AUC=.674 [.481, .748], *p*=.006; PRS-C: AUC=.819 [.718, .920], p<.001). Figure 1 provides a visual depiction of the ROC curve and AUC for the analysis differentiating between children with and without ADHD for the TRS-C and PRS-C.

As the TRS-C and PRS-C ROC analyses, respectively, resulted in good and excellent AUCs when classifying children with ADHD, a cutoff score was calculated based on sensitivity and specificity values. Based on a Youden’s J statistic for the TRS-C, a cutoff score of 51 showed the highest sum of sensitivity (.785) and specificity (.714). For the PRS-C, a cutoff score of 32 showed the optimal level of sensitivity (.908) and specificity (.844). A cutoff score was not calculated for the post-hoc analysis, as this analysis served to demonstrate the utility of the screener at detecting a specific diagnosis related to executive dysfunction (i.e., ADHD) rather than provide diagnostic guidance when distinguishing between ADHD and RD.

Discussion

Complementing previous results on the BASC and BASC-2-EF (Garcia-Barrera et al., 2011, 2013, 2015; Karr & Garcia-Barrera, 2017; Sadeh et al., 2012), the current findings offer further psychometric evidence for the four-factor screener, and new evidence for the validity of the GSS at differentiating between typically developing children and those with ADHD. While the screener has been replicated multiple times using the original BASC (see Garcia-Barrera et al., 2014), this study is the first to replicate the screener using the BASC-2 with a mixed clinical sample, and the first study to offer evidence for a second-order executive function factor. The good fit for the second-order factor model rationalized the calculation of a parsimonious GSS, which was used in the ROC curve analyses. The results of these analyses met standards for good and excellent AUC values for the TRS-C and PRS-C, respectively, when differentiating ADHD from typical development, evidencing strong sensitivity and specificity of the BASC-2-EF GSS.

Although both the teacher and parent ratings resulted in significant AUC values, there were slight differences in the magnitude of AUC values between the two raters across all ROC curve analyses. These discrepancies could result from the diagnostic process for ADHD, where a clinical interview and DSM-IV questionnaire completed with the parent were the primary sources of information to determine a diagnosis. They also could result from both the biases of raters and the differences in settings in which the raters interact with children (Narad et al., 2015). For ADHD assessment, both raters had highly significant AUC values; however, when comparing children with and without RD, the teacher ratings showed greater executive dysfunction among the RD children, whereas the parent ratings did not. Within a classroom setting, a child with RD may more quickly disengage or become frustrated with a reading task, which may be perceived as reduced attentional or emotional control, respectively. In contrast, parent raters may provide more one-on-one support to their children when reading at home, and, thus, may not observe as many of these behaviors. Differences in parent and teacher ratings have been historically evident, such as more inconsistent and biased ratings provided by parents than teachers (Hartman, Rhee, Willcutt, & Pennington, 2007; Lochman, 1995). Lower reliability of parent ratings also was observed in the current findings; however, parent ratings evidenced greater validity as a diagnostic measure of ADHD based on AUC values.

These differences in parent and teacher ratings made evident the value of a multi-informant assessment process. A teacher may provide a more reliable rating of executive functions, whereas a parent may provide a rating with better diagnostic utility for ADHD. Parents have been shown to have better diagnostic determination for ADHD than teachers (despite bias and reliability issues), although utilizing a composite from multiple raters is better diagnostically than relying on one rater (O'Neill, Schneiderman, Rajendran, Marks, & Halperin, 2014; Willcutt, 2012). Similarly, neither the parent nor teacher ratings on the BASC-2-EF are independently diagnostic of ADHD, and future research should evaluate the incremental validity of the BASC-2-EF scales to examine their added contribution to ADHD diagnosis beyond neuropsychological measures or other diagnostic scales. As opposed to low performance on neuropsychological measures, informant-reports of impaired executive functions may result from perceived problems by the rater that are unrelated to executive functions, in addition to true executive function deficits. This is evident in the current findings, where teacher ratings on the BASC-2-EF differentiated between RD and typical development. Although some evidence has linked RD with deficits in executive function (e.g., Willcutt et al., 2001), executive dysfunction would not be considered a primary diagnostic feature of RD. The behavioral measure cannot replace a close examination of cognitive/linguistic functioning and reading abilities in determining an RD diagnosis.

As opposed to RD, far more evidence has demonstrated pronounced executive function deficits among children with ADHD (Bloom, Miller, Garcia, & Hynd, 2005; Willcutt et al., 2005), and this executive function screener effectively differentiated ADHD from RD and controls. Nonetheless, although there has been an increased appreciation for the ecological validity of behavioral ratings of executive function (Garcia-Barrera et al., 2011), they do not necessarily correspond to deficits found on testing, considering that correlations between cognitive and behavioral measures often present as non-significant (Toplak et al., 2013). The behavioral assessment of executive functions has become increasingly popular in both research (e.g., Barkley & Fischer, 2011; Sadeh et al., 2012) and clinical practice (Rabin et al., 2016), but it does not replace a comprehensive neuropsychological assessment that closely evaluates cognitive deficits that are detectable through performance-based measures. Other neuropsychological measurements have demonstrated sensitivity and specificity in ADHD assessment (Holmes et al., 2010), and the BASC-2-EF can be coupled with such information to create a clearer diagnostic picture, based on both cognitive measurement and behavioral ratings. The current study is limited by the inclusion of the BASC-2 Attention Problems and Hyperactivity subscales as one part of the diagnostic process for ADHD, as the items from both subscales overlap with the BASC-2-EF. Further research on the incremental validity of the BASC-2-EF at detecting ADHD beyond preexisting subscales is warranted. Research on diagnostic cutoffs of the executive function screener using the BASC-3 also would be useful for direct implementation into current clinical practice, whereas the current findings will more likely support diagnostic decision making in archival datasets.

References

Barkley, R. A., & Fischer, M. (2011). Predicting impairment in major life activities and occupational functioning in hyperactive children as adults: Self-reported executive function (EF) deficits versus EF tests. *Developmental Neuropsychology*, *36*(2), 137-161.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588-606.

Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., ... & Faraone, S. V. (2004). Impact of executive function deficits and attention-deficit/hyperactivity disorder (ADHD) on academic outcomes in children. *Journal of Consulting and Clinical Psychology*, *72*(5), 757-766.

Bloom, J. S., Miller, C. J., Garcia, M. A., & Hynd, G. W. (2005). Reading disabilities in children with attention deficit hyperactivity disorder. In D. Gozal & D. L. Molfiese (Eds.), *Attention Deficit Hyperactivity Disorder* (pp. 337-358). Totowa, NJ: Humana Press.

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

Duggan, E. C., Garcia-Barrera, M. A., & Müller, U. (2016). Derivation, replication, and validity analyses of a screener for the behavioral assessment of executive functions in young adults. *Assessment*. Advance online publication. doi:10.1177/1073191116670511

Garcia-Barrera, M. A., & Kamphaus, R. W. (2006). Attention Deficit Hyperactivity Disorder. In R. W. Kamphaus & J. Campbell (Eds), *Psychodiagnostic Assessment of Children: Dimensional and Categorical Methods* (pp. 319-355). New York: Wiley Press.

Garcia-Barrera, M. A., Duggan, E. C., Karr, J. E., & Reynolds, C. R. (2014). Examining executive functions using the Behavior Assessment System for Children. In S. Goldstein

& J. Naglieri (Eds.) *Handbook of executive functioning* (pp. 283-299). New York: Springer Publishing Co.

Garcia-Barrera, M. A., Kamphaus, R. W., & Bandalos, D. (2011). Theoretical and statistical derivation of a screener for the behavioral assessment of executive functions in children. *Psychological Assessment, 23*, 64–79.

Garcia-Barrera, M. A., Karr, J. E., Duran, V., Direnfeld, E. Y., & Pineda, D. A. (2015). Cross-cultural validation of a behavioral screener for executive functions: Guidelines for clinical use among Colombian children with and without ADHD. *Psychological Assessment*, *27*(4), 1349-1363.

Garcia-Barrera, M. A., Karr, J. E., & Kamphaus, R. W. (2013). Longitudinal applications of a behavioral screener of executive functioning: Assessing factorial invariance and exploring latent growth. *Psychological Assessment, 25*(4), 1300-1313.

Graziano, P. A., McNamara, J. P., Geffken, G. R., & Reid, A. M. (2013). Differentiating co-occurring behavior problems in children with ADHD: Patterns of emotional reactivity and executive functioning. *Journal of Attention Disorders*, *17*(3), 249-260.

Hartman, C. A., Rhee, S. H., Willcutt, E. G., & Pennington, B. F. (2007). Modeling rater disagreement for ADHD: are parents or teachers biased? *Journal of Abnormal Child Psychology*, *35*(4), 536-542.

Holmes, J., Gathercole, S. E., Place, M., Alloway, T. P., Elliott, J. G., & Hilton, K. A. (2010). The diagnostic utility of executive function assessments in the identification of ADHD in children. *Child and Adolescent Mental Health*, *15*(1), 37-43.

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.

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*(2), 83-93.

Karr, J. E., & Garcia-Barrera, M. A. (2017). The assessment of executive functions using the BASC-2. *Psychological Assessment, 29*(9), 1182-1187.

Kibby, M.Y., Dyer, S.M., Vadnais, S.A., Jagger, A.C., Casher, G.A., & Stacy, M. (2015). Visual processing in reading disorders and attention-deficit/hyperactivity disorder and its contribution to basic reading ability. *Frontiers in Psychology, 6*(1635), 1-11.

Kibby, M.Y., Lee, S.E., & Dyer, S.M. (2014). Reading performance is predicted by more than phonological processing. *Frontiers in Psychology*, *5*(960), 1-7.

Lochman, J. E. (1995). Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology, 63,* 549-559.

Mahan, S., & Matson, J. L. (2011). Children and adolescents with autism spectrum disorders compared to typically developing controls on the Behavioral Assessment System for Children, (BASC-2). *Research in Autism Spectrum Disorders, 5*(1), 119-125.

Muthén, L, & Muthén, B. (2015). MPlus (Version 7.4) [Computer Software]. Los Angeles: Muthén & Muthén.

Narad, N., Garner, A., Peugh, J., Tamm, L., Antonini, T., Kingery, K., Simon, J. & Epstine, J. (2015). Parent-Teacher Agreement on ADHD Symptoms Across Development. *Psychological Assessment*, *27*(1), 239-248.

O'Neill, S., Schneiderman, R. L., Rajendran, K., Marks, D. J., & Halperin, J. M. (2014). Reliable ratings or reading tea leaves: Can parent, teacher, and clinician behavioral ratings of preschoolers predict ADHD at age six? *Journal of Abnormal Child Psychology*, *42*(4), 623-634.

Pineda, D. A., Aguirre, D. C., Garcia, M. A., Lopera, F. J., Palacio, L. G., & Kamphaus, R. W. (2005). Validation of two rating scales for attention-deficit hyperactivity disorder diagnosis in Colombian children. *Pediatric Neurology*, *33*(1), 15-25.

Pineda, D. A., Puerta, I. C., Aguirre, D. C., García-Barrera, M. A., & Kamphaus, R. W. (2007). The role of neuropsychologic tests in the diagnosis of attention deficit hyperactivity disorder. *Pediatric Neurology*, *36*(6), 373-381.

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*(3), 206-230.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior assessment system for children* (3rd ed.). Bloomington, MN: Pearson.

Sadeh, S. S., Burns, M. K., & Sullivan, A. L. (2012). Examining an executive function rating scale as a predictor of achievement in children at risk for behavior problems. *School Psychology Quarterly, 27*, 236-246.

Steiger, J. H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH*. Evanston, IL: SYSTAT.

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.

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131-143.

Willcutt, E. G. (2012). The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics, 9(3), 490-499.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry*, *57*(11), 1336-1346.

Willcutt, E. G., Pennington, B. F., Boada, R., Ogline, J. S., Tunick, R. A., Chhabildas, N. A., & Olson, R. K. (2001). A comparison of the cognitive deficits in reading disability and attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology*, *110*(1), 157-172.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, *3*(1), 32-35.

Table 1. *Standardized Model Results*

BASC-2-EF: TRS-C					BASC-2-EF: PRS-C				
2nd-Order Factor	1st-Order Factor	λ	Item	λ	2nd-Order Factor	1st-Order Factor	λ	Item	λ
EF	PS	0.872	i30*	0.746	EF	PS	0.839	i4*	0.632
			i50	0.614				i36*	0.756
			i58*	0.583				i67*	0.832
			i73*	0.896				i113*	0.785
			i78*	0.778				i132*	0.714
			i86*	0.625				i154*	0.523
			i88*	0.543		AC	0.931	i9	0.838
			i91*	0.908				i17*	0.890
			i117*	0.615				i41*	0.881
	AC	0.937	i5	0.835				i49*	0.739
			i33*	0.89				i73	0.814
			i44*	0.897				i105*	0.927
			i61	0.875		BC	0.739	i38	0.613
			i72*	0.874				i52	0.813
			i100	0.897				i116	0.861
			i128*	0.85				i148	0.836
	BC	0.794	i8	0.744		EC	0.733	i10	0.596
			i10	0.783				i14*	0.644
			i14	0.786				i46*	0.714
			i18	0.927				i90	0.647
			i38	0.905					
			i46	0.881					
			i54	0.840					
			i66	0.906					
			i74	0.700					
			i94	0.761					
			i102	0.927					
			i120	0.694					
	EC	0.796	i29	0.604					
			i36	0.854					
			i49	0.724					
			i57*	0.747					
			i85*	0.678					

Note. *Indicates an item that was reverse coded before summing (i.e., 1=4, 2=3, 3=2, 4=1). AC = Attentional Control; BC = Behavioral Control; EC = Emotional Control; EF = Executive Function; PS = Problem Solving.

Table 2. *Model Building Analysis*

Model: TRS-C	WLMSV χ^2 (*p*)	*df*	CFI	ΔCFI	TLI	RMSEA (90% C.I.)
Model 1: 33 Items, 1 Factor	2125.119 (.0000)	495	.888		.881	.121 (.116-.126)
Model 2: 33 Items, 2 Factors	1946.191 (.0000)	494	.900	.012	.893	.114 (.109-.120)
Model 3: 33 Items, 3 Factors	1433.954 (.0000)	492	.935	.035	.931	.092 (.087-.098)
Model 4: 33 Items, 4 Factors	1330.278 (.0000)	489	.942	.007	.938	.087 (.082-.093)
Model 5: Second-Order Factor	1442.673 (.0000)	491	.935	-.007	.930	.093 (.087-.098)
Model: PRS-C	WLMSV χ^2 (*p*)	*df*	CFI	ΔCFI	TLI	RMSEA (90% C.I.)
Model 1: 20 items, 1 Factor	734.086 (.0000)	170	.897		.885	.125 (.116-.134)
Model 2: 20 items, 2 Factors	631.344 (.0000)	169	.916	.019	.905	.114 (.104-.123)
Model 3: 20 items, 3 Factors	435.108 (.0000)	167	.951	.035	.945	.087 (.077-.097)
Model 4: 20 items, 4 Factors	385.839 (.0000)	164	.960	.009	.953	.080 (.070-.090)
Model 5: Second-Order Factor	454.388 (.0000)	166	.948	-.012	.940	.091 (.081-.101)

Figure 1. *Receiver Operating Characteristic (ROC) curve for the Behavior Assessment System for Children, Second Edition, Executive Function screener (BASC-2-EF) Global Sum Score using the Teacher Rating Scale for Children (TRS-C; Left) and Parent Rating Scale for Children (PRS-C; Right).*