Please use the following citation when referencing this work:

McGill, R. J. (in press). Test review of Behavior Rating Inventory of Executive Function, Second Edition. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), The twenty-first mental measurements yearbook (pp. xx-xx). Lincoln, NE: Buros Center for Testing.

Review of the Behavior Rating Inventory of Executive Function, Second Edition by Ryan J. McGill, Assistant Professor of School Psychology, School of Education, College of William & Mary, Williamsburg, VA:

Description

The Behavior Rating Inventory of Executive Function, Second Edition (BRIEF2) is the first revision of the BRIEF (Gioia, Isquith, Guy, & Kenworthy, 2000), the first published rating scale measuring self-regulatory abilities in children and adolescents. The BRIEF2 maintains many of the essential properties of the BRIEF with numerous scales designed to provide "a window into the everyday behavior associated with specific domains of self-regulated cognitive problem solving, behavioral functioning, and emotional regulation" (Gioia, Isquith, Guy, & Kenworthy, 2015, p. 26). The BRIEF family of instruments relies on a bi-level model of executive functioning in which specific functions are embedded within a broader network of cognitive/affective domains. Nevertheless, the BRIEF2 is a substantive revision and re-standardization with a larger and more representative normative sample and more concise scales. Separate co-normed parent, teacher (63 items each), and self-report (55 items) forms are available.

The BRIEF2 parent and teacher forms include nine clinical scales (Inhibit, Self-Monitor, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Task-Monitor, Organization of Materials) which combine to form three first-order index scores (Behavioral Regulation Index [BRI], Emotional Regulation Index [ERI], Cognitive Regulation Index [CRI]) and an omnibus

second-order Global Executive Composite (GEC) score. The self-report form yields all of the same index and composite scores as the parent and teacher forms with a slightly different configuration of seven clinical scales. Additionally, all three forms contain several validity indices (negativity, infrequency, inconsistency) to help clinicians identify questionable responding by raters.

Development

The goal for the revision noted in the Professional Manual was to enhance the clinical and research utility of the scale without substantively altering the structure of the instrument. Specific enhancements include improving psychometric properties with new statistics (i.e., change indices, base rate tables) to support clinical interpretation, the addition of an infrequency scale to help identify unusual responding, and greater parallelism in item content and order across the forms. It should also be noted that no new items were added as part of the revision although some items were deleted to make the scales more concise and help decrease the frequency of nonresponse by raters.

The BRIEF 2 was standardized and validated for use with boys and girls aged 5-18 years (parent and teacher forms) and boys and girls aged 11-18 years (self-report form). The Manual indicates that test items were written at approximately a fifth-grade reading level and each of the forms can be completed in less than 10 minutes. Rating forms are accompanied by a Professional Manual and a Fast Guide which contain standardized oral instructions for raters and several case studies to aide users with clinical interpretation. A supplemental Interpretive Guide (Isquith, Gioia, Guy, & Kenworthy, 2017) is also available containing additional case studies and intervention recommendations.

Technical

Standardization

The BRIEF2 Professional Manual presents extensive and detailed information on the normative sample. Sample sizes ranged from 803 (self-report form) to 1,400 (parent and teacher forms) participants. The standardization sample was obtained through stratified proportional sampling across key demographic variables (age, gender, race/ethnicity, parent educational level as a proxy for socioeconomic status, and geographic region). Inspection of the demographic tables reported in the Manual reveal a close match to the 2013 U.S. Census data.

Separate norms are provided for boys and girls across four age groups. Inspection of the raw score conversion tables provided in Manual revealed the presence of item gradient violations for several of the clinical scales. For example, the average gradient for the Initiate scale on the teacher rating form for girls aged 11-13 years was 4.2 points exceeding established guidelines (Bracken, 1987). According to Wasserman and Bracken (2013), inadequate item density across the distribution of a latent trait reduces the sensitivity and discrimination of a test. Whereas the ceilings for all scales and index scores were adequate, virtually all scales suffered from floor effects. Although it should be noted that this is less of a concern for rating scales such as the BRIEF2 in which lower scores are indicative of intact abilities.

All score scale, index, and composite scores on the BRIEF2 are expressed as *T* scores ($M = 50$, $SD = 10$, range = 35 to ≥ 90), though users should be aware that the percentile ranks associated with those scores may vary from scale to scale due to skew in the distributions. According to the Manual, *T* scores between 60 and 64 are considered to be *mildly elevated,* scores between 65 and 69 are considered *potentially clinically elevated,* and scores at or above 70 are considered *clinically elevated* and significant.

In terms of interpretation, clinicians are encouraged to utilize BRIEF2 scores to develop hypotheses about individual strengths and weaknesses that can be explored in follow-up testing. To wit, a series of procedures are described that includes assessing the validity indices, interpreting clinical scale and index scores, interpreting profile shape and scatter, and comparing the results obtained across multiple informants. Like many psychoeducational instruments, it is suggested that the GEC score can only be validly interpreted if an examiner first determines that there are no significant differences among the index scores though the use of this popular interpretive heuristic has been questioned (McGill, 2016).

Reliability

Three methods of estimating reliability of BRIEF2 scores are reported in the Professional Manual: internal consistency, test-retest stability, and interscorer agreement. Internal consistency was estimated using Cronbach's alpha. Average coefficients across the forms were moderate to strong ranging from .81 to .98 and estimates were invariant across several demographic characteristics. Temporal stability was assessed using a subsample of the normative sample with retest intervals that ranged from 2.9 to 3.7 weeks. Average test-retest correlations for the parent (.67 to .92), teacher (.76 to .90), and self-report (.61 to .85) forms were moderate to strong. As expected, interrater reliability coefficients ranged from low to moderate on all of the rating forms although higher levels of agreement were observed for the scores obtained from parent dyads.

Validity

Consistent with the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014), evidence for validity was structured around the areas of test content, internal structure, and relations with other variables. To examine the structure of the

BRIEF2, preference for confirmatory factor analysis (CFA) was argued and evidence for internal structure relied exclusively on the use of constrained CFA analyses (i.e., rival models were not examined). The first CFA examined relationships at the item-level in order to establish the latent structure of the clinical scales. Then a second CFA commenced using the clinical scale scores as measured variables to examine higher-order structure using a three oblique (correlated) factors model. Results from both analyses generally supported the theoretical model posited in the Manual for all of the rating forms.

Numerous comparative studies of the BRIEF2 in relation to other behavioral rating scales and intelligence tests were reported in the Manual, and evidence for convergent and divergent validity was inconsistent. Whereas the zero-order correlations between scale and index scores from the BRIEF2 and other attention-related indices on rating scales were moderate to strong, relationships with some reference scores on intelligence tests was less robust. For example, correlations between the Working Memory scale on the BRIEF2 and the Working Memory Index on the WISC-IV/WAIS-IV were -.08 and -.27 respectively for the parent and teacher forms. Relationships between BRIEF2 scores and external measures of achievement were not reported. Special groups and matched controls were compared to test for specific group differences. Mean score differences between the groups were typically what one would expect with individuals diagnosed with various subtypes of attention-deficit/hyperactivity disorder, learning disabilities, and autism spectrum disorder scoring higher than matched groups of typically developing peers. To buttress these findings, additional diagnostic efficiency statistics were provided to support the use of specific clinical scales for differential diagnosis.

Commentary

As with any test, the BRIEF2 has strengths and weaknesses. Whereas extensive psychometric information is provided in the Manual, there are a few notable omissions. In particular, it remains unclear to this reviewer how the instrument should be interpreted. Given the strong correlations and linear redundancy between the scale and index scores, it seems relevant to ask whether sufficient variance is captured by the all of the scores that are presented to clinicians as capable of being interpreted. In fact, the GEC score was not included in any of the structural models that were examined even though some recent CFA studies (e.g., Roth et al., 2013) have supported a three-stratum model with the GEC score at the apex. As previously noted, strong latent factor correlations were observed in all of the first-order models that were reported in the Manual which imply the presence of a higher-order dimension (Gorsuch, 1983). Unfortunately, this dimension was not explored. As noted by Reise, Haviland, and Moore (2010), data from behavioral rating scales are rarely univocal, often containing variance that can be sourced to multiple latent dimensions and attributes producing interpretive ambiguity. Given that users of the BRIEF2 are encouraged to use the clinical scales as a primary means to conduct profile analysis, it is important to determine how much unique information is captured by those scores apart from a more general construct that may explain variance in all of the BRIEF2 measures. This information is vital given the well documented issues with psychoeducational profile analysis in general (Youngstrom, 2013).

Additionally, information regarding the relationship between BRIEF2 scores and performance measures of executive functioning (e.g., Trail Making and n-back tasks) is absent. This omission is curious given the fact that the BRIEF was originally developed as “an index of ecological validity for findings on performance measures of executive function in the clinical setting” (Isquith, et al., 2017, p. 21). In particular, this information would be useful for helping

clinicians to select appropriate measures for follow-up testing and other useful procedures for validating the inferences generated from BRIEF2 data.

Summary

Overall, the BRIEF2 has many strengths. The test materials are of exceptional quality and the rating forms are easy to administer and score. The Professional Manual contains a wealth of psychometric information and for the most part, the technical properties of the instrument are quite exceptional. The use of a larger and more representative normative sample addresses one of the more prominent issues noted in previous reviews of the BRIEF and the use of diagnostic efficiency statistics to help inform clinical utility is noteworthy. In the Manual it is made clear that the BRIEF2 should not be used as a standalone measure for diagnosing executive dysfunction. Users who heed this advice and interpret BRIEF2 scores with appropriate caution within the context of a comprehensive assessment battery will likely find the instrument to be a useful addition to their clinical toolkit.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement on Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5,* 313-326. doi: 10.1177/073428298700500402

Carretta, T. R., & Ree, J. J. (2001). Pitfalls of ability research. *International Journal of Selection and Assessment, 9,* 325-335. doi: 10.1111/1468-2389.00184

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function*. Lutz, FL: PAR.

Gioia, G. A., Isquth, P. K., Guy, S. C., & Kenworthy, L. (2015). *Behavior Rating Inventory of Executive Function, Second Edition professional manual*. Lutz, FL, PAR.

Gioia, G. A., Isquth, P. K., Guy, S. C., & Kenworthy, L. (2017). *Behavior Rating Inventory of Executive Function, Second Edition interpretive guide*. Lutz, FL, PAR.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.

McGill, R. J. (2016). Invalidating the full scale IQ score in the presence of significant factor score variability: Clinical acumen or clinical illusion? *Archives of Assessment Psychology, 6 (1),* 49-79. Retrieved from http://www.assessmentpsychologyboard.org/journal/index.php/AAP

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

Roth, R. M., Lance, C. E., Isquith, P. K., Fisher, A. S., & Giancola, P. R. (2013). Confirmatory factor analysis of the Behavior Rating Inventory of Executive Function-Adult Version in healthy adults and application to attention-deficit/hyperactivity disorder. *Archives of Clinical Neuropsychology, 28,* 425-434. doi: 10.1093/arclin/act031

Waserman, J. D., & Bracken, B. A. (2013). Fundamental psychometric considerations in assessment. In I. B. Weiner, J. R. Graham, & J. A. Naglieri (Eds.). *Handbook of psychology: Assessment psychology* (2nd ed., Vol. 10, pp. 50-81). Hoboken, NJ: John Wiley.

Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child and Adolescent Psychology, 42,* 139-159. doi: 10.1080/15374416.2012.736358