American Journal on Intellectual and Developmental Disabilities

Implications of using the BRIEF-preschool with school-age children with Down syndrome

--Manuscript Draft--

Manuscript Number:	AJIDD-D-22-00041R3
Article Type:	Research Report
Keywords:	Down Syndrome; trisomy 21; executive functioning; BRIEF-2; BRIEF-P
Corresponding Author:	Anna J Esbensen, PhD Cincinnati Children's Hospital Medical Center Cincinnati, OH UNITED STATES
First Author:	Anna J Esbensen, PhD
Order of Authors:	Anna J Esbensen, PhD
	Emily K Schworer
	Nancy R Lee
	Emily K Hoffman
	Kaila Yamamoto
	Deborah Fidler
Manuscript Region of Origin:	UNITED STATES
Abstract:	This study evaluated the appropriateness of scoring the BRIEF-preschool (BRIEF-P) using age-equivalent scores generated from multiple measures of cognition and language among school-age children with Down syndrome (DS). Subscale T-scores for 95 children with DS were contrasted using standard scoring on the BRIEF-2 (based on chronological age) to alternate scoring using the BRIEF-P (based on age-equivalent) for patterns of subscale intercorrelations, differences in mean scores, and agreement on findings from clinical cut-off scores. Results with children with DS suggested using (1) the BRIEF-P for children ages 2-5 years old, (2) the BRIEF-2 with chronological-age scoring or the BRIEF-P with age-equivalent scoring (with some caveats) for research on children ages 5-10 years old, and (3) the BRIEF-2 for children ages 11 and older.

Implications of using the BRIEF-preschool with school-age children with Down syndrome

Abstract

This study evaluated the appropriateness of scoring the BRIEF-preschool (BRIEF-P) using age-equivalent scores generated from multiple measures of cognition and language among school-age children with Down syndrome (DS). Subscale T-scores for 95 children with DS were contrasted using standard scoring on the BRIEF-2 (based on chronological age) to alternate scoring using the BRIEF-P (based on age-equivalent) for patterns of subscale intercorrelations, differences in mean scores, and agreement on findings from clinical cut-off scores. Results with children with DS suggested using (1) the BRIEF-P for children ages 2-5 years old, (2) the BRIEF-2 with chronological-age scoring *or* the BRIEF-P with age-equivalent scoring (with some caveats) for research on children ages 5-10 years old, and (3) the BRIEF-2 for children ages 11 and older.

Keywords: Down syndrome, trisomy 21, executive functioning, BRIEF-2, BRIEF-P

Implications of using the BRIEF-preschool with school-age children with Down syndrome

Individuals with Down syndrome (DS) frequently present with difficulties in the areas of executive functioning (Tungate & Conners, 2021). Executive functions, a set of cognitive regulation skills, are responsible for allowing one to manage, coordinate, organize, assemble, order, and monitor multiple tasks (Diamond, 2013). They include skills related to resisting impulses, shifting between tasks, regulating emotions, mentally manipulating information (working memory), and planning and organizing materials. Difficulties in these areas of executive functioning in DS are present in comparison to chronological-age-matched and mental-age-matched typically developing peers (Camp et al., 2016; Daunhauer et al., 2014; Daunhauer et al., 2017; Lanfranchi et al., 2010; Lanfranchi et al., 2009; Vicari et al., 1995). Difficulties are commonly reported in the ability to transition between tasks and use working memory, with some mixed findings related to the ability to inhibit responses (Camp et al., 2016; Daunhauer et al., 2014; Daunhauer et al., 2017; Esbensen et al., 2019; Lee et al., 2011; Pennington et al., 2003). Further, a distinct pattern of strengths and weaknesses in areas of executive function is present across preschool and school-age children with DS, specifically with strengths in emotional control and weaknesses in working memory relative to their chronological-age peers (Onnivello et al., 2021). Executive functioning plays a key role in adaptive, academic, behavioral, and language outcomes for children with DS (Baddeley & Jarrold, 2007; Esbensen et al., 2021; Will et al., 2017; Will et al., 2021).

Challenges in executive functioning for individuals with DS are identified in both neuropsychological tasks and in parent- and teacher-report forms (Daunhauer et al., 2014; Daunhauer et al., 2017; Esbensen et al., 2019; Lee et al., 2011; Loveall et al., 2017; Will et al.,

2017). Yet there are inconsistent patterns of strengths and weaknesses among subdomains of executive functioning reported among children with DS. These inconsistent findings are typically found between laboratory-based measures versus parent-report measures, for example with shifting being reported as an impaired subdomain when using laboratory measures and also a relative strength by parents and teachers compared to mental age-matched typically developing peers (Daunhauer et al., 2014; Lee et al., 2011; Loveall et al., 2017; Tungate & Conners, 2021), and also between verbal and visuospatial task modality (Kittler et al., 2006; Rowe et al., 2006).

Given these divergent findings on challenges in laboratory-based measures of executive functioning when compared to typically-developed peers, recent efforts have evaluated different neuropsychological measures of executive functioning in order to recommend measures appropriate for assessing this construct among individuals with DS (d'Ardhuy et al., 2015; Edgin et al., 2017; Lanfranchi et al., 2004; Lanfranchi et al., 2010; Lanfranchi et al., 2009; Schworer et al., 2022). These efforts have also included evaluation of parent- and teacher-report forms of executive functioning among individuals with DS (Esbensen et al., 2019; Onnivello et al., 2021). Specifically, the school-age version of the Behavior Rating Inventory of Executive Function (BRIEF) has sound psychometric properties when administered to school-age children with DS (Esbensen et al., 2019). The BRIEF and its second edition, the BRIEF-2, both intended for children ages 5-18 years old, are commonly used in studies of school-age children with DS (Esbensen et al., 2021; Esbensen et al., 2017; Lee et al., 2015; Loveall et al., 2017; Memisevic & Sinanovic, 2014; Onnivello et al., 2021). The preschool version of the BRIEF (BRIEF-P), intended for children ages 2-5 years old, is also commonly used when assessing preschoolers with DS (Loveall et al., 2017; Onnivello et al., 2021). In these studies, the version of the BRIEF that was administered corresponded with the chronological age of the child with DS. This

approach has the benefit of scoring children based on the normative data of the version of the BRIEF used and demonstrating where children present with strengths in executive functioning, relative to the typically developing normative group. However, this approach may assess weaknesses in executive functioning that are related to the child's overall developmental status, particularly among older children with DS who may have larger gaps in their skills than younger children, compared to typically developing peers (DiStefano et al., 2020). This approach becomes problematic when the research question pertains to measuring challenges with executive functioning beyond the child's overall developmental status. Further, item wording on the BRIEF-2 may be age-appropriate, but not relevant (i.e., homework) or not developmentally relevant to school-age children with DS if others support the child with performing that skill.

Researchers have also used the BRIEF-P in studies of children with DS above the intended chronological age (over 5 years old) with assessments scored based on developmental level (d'Ardhuy et al., 2015; Daunhauer et al., 2014; Lee et al., 2011; Pritchard et al., 2015; Schworer et al., 2019). This approach has the benefit of comparing individuals with DS to mental-age-matched typically developing peers and interpreting findings relative to the individual's developmental level. Thus, a child's score would reflect their performance in executive functioning that accounts for their overall cognitive ability. Another advantage of this approach is that the age range of the individuals with DS is not limited to the particular BRIEF version, allowing for lifespan approaches or broader research questions to be examined (Lee et al., 2011; Pritchard et al., 2015). Use of the BRIEF-P restricted to school-age children is also common in the research literature (Daunhauer et al., 2014; Schworer et al., 2019). However, similar to concerns with item wording on the BRIEF-2 being less pertinent for school-age children with DS, the items on the BRIEF-P may also be less pertinent to the chronological age

of the older individuals with DS being rated. Items on the BRIEF-P refer to playmates and engaging in tasks such as pretend play, which are difficult for caregivers to rate with older children, adolescents, or young adults. When the BRIEF-P was evaluated for use among adults with DS, adequate test-retest reliability was demonstrated, yet the findings suggested that its use among adults was less appropriate than for adolescents with DS (d'Ardhuy et al., 2015).

The implications of scoring the BRIEF-P when used with school-age children with DS has not been evaluated. Understanding the effects of using the BRIEF-P with school-age children has importance both for research and clinical practice. These implications are particularly pertinent to 5-6-year-olds, for whom either the BRIEF-P or BRIEF-2 can be administered. Adding to the complexity, when the BRIEF-P is used for school-age children, scoring procedures vary across studies for participants out of the age range of the measure. Some studies have used raw scores (d'Ardhuy et al., 2015; Schworer et al., 2019). Alternatively, an age-equivalent score generated from other standardized tests of cognition or language is used for scoring with normative data. Age-equivalent scores used for scoring the BRIEF-P are predominantly from nonverbal cognitive skills, such as the Leiter-R (Daunhauer et al., 2014) and Leiter-3 (d'Ardhuy et al., 2015); but see others have used the average of DAS/MSEL verbal and nonverbal subtest scores (Lee et al., 2011). Age-equivalents from nonverbal cognitive measures are commonly used as they exclude the weakness in expressive language when generating a T-score. Use of an age-equivalent based on a standardized test of cognition (involving both verbal and nonverbal skills) may be lower than one obtained from a standardized test of receptive language yet may also be more representative of a child's overall developmental level. However, it is unclear exactly how age-equivalent scores from various assessment measures differ, or how they may impact T-scores generated on the BRIEF-P. Guidance is needed for which assessments to use to

generate age-equivalent scores for scoring the BRIEF-P, and for whether using these age-equivalent scores to generate T-scores on the BRIEF-P with school-age children with DS is appropriate for research or clinical practice.

This study included both the BRIEF-2 and BRIEF-P parent forms to investigate the appropriateness of the BRIEF-P for use with school-age children with DS. The aim was to compare different scoring procedures to standardize recommendations for use of the BRIEF in both research and clinical practice. Our first goal was to describe the appropriateness of age-equivalent scores generated by different standardized tests of cognition and language to the age range of the BRIEF-P. Second, we aimed to understand the implications for BRIEF-P scoring when using age-equivalent scores generated by different measures of cognition/language for *research*, with a focus on implications of T-scores generated by the developmental-age scoring on the BRIEF-P in comparison to T-scores generated by the chronological-age scoring on the BRIEF-2. And third, we examined the implications for BRIEF-P scoring when using age-equivalent scores generated by different measures of cognition/language for *clinical practice*, with a focus on agreement in decisions based on clinical cut-off scores generated by the BRIEF-P or BRIEF-2.

Method

Participants

Participants consisted of 95 children with DS and their parents. Children ranged in chronological age from 6 to 17 years (M = 12.53, SD = 3.26), reflecting the inclusion criteria for a larger study. Children were distributed evenly across sex (48.4% female), yet were predominantly White (87.4%; 5.3% Black, 4.2% Asian, 3.2% other race) and non-Hispanic

(93.7%). Children had a documented diagnosis of DS and the primary language at home was English. Parental respondents were primarily mothers (93.7%).

Procedures

Data were extracted from participants in a larger multi-site longitudinal study evaluating outcome measures among children with DS. Families were recruited through local DS associations, a pediatric hospital, and through a DS specialty clinic. Inclusion criteria for the larger study were: having a diagnosis of DS (Trisomy 21, translocation, or mosaicism), being between the ages of 6-17 years, and having English be the primary language spoken at home. In addition to these inclusionary criteria, children included in the current analyses must also have had complete data on the BRIEF-P and BRIEF-2.

Four standardized tests of cognition or language were administered to the child at a baseline visit as part of the larger study design to generate age-equivalent scores. In a quiet testing room, children were administered the Kaufman Brief Intelligence Test – second edition (KBIT-2), the abbreviated version of the Stanford-Binet Intelligence Scales, fifth edition (SB-5), the Peabody Picture Vocabulary Test, fifth edition (PPVT-5), and the Expressive Vocabulary Test, third edition (EVT-3) (Dunn, 2019; Kaufman, 2004; Roid, 2003; Williams, 2018). All four measures are considered appropriate for children and adults with DS (Edgin et al., 2010; Esbensen et al., 2017). Assessments of cognition and language were completed in a structured order, as listed above, to facilitate engagement with the child between nonverbal and verbal tasks. Alternative start points were used. All children started at the first item on the KBIT-2. Clinical judgment was used for the start point on the SB-5. Raw score performance on the SB-5 Fluid Reasoning was used to create standardized alternative start points for administration of the PPVT-5, and on the SB-5 Vocabulary to create standardized alternative start points for

administration of the EVT-3. Similar to how the SB-5 routing tasks guide administration of the nonverbal and verbal portions of the SB-5, alternative start points ensured that children were started on the PPVT-5 and EVT-3 consistent with their level of performance on the SB-5. Reverse rules were used if the alternative start points were too high. Total testing time generally ranged from 45-60 minutes. Children were provided visual schedules, and snack or small toy incentives at the end of each visit.

Parents provided demographic information about their child with DS and completed parent rating forms on their child's executive functioning, using standard administration instructions. All study activities were approved and overseen by the Institutional Review Board (IRB), using Single IRB at the supervising medical center providing oversight for the multisite study. Families were paid for their participation.

Measures

Cognition and Language

The KBIT-2 is a brief assessment of cognitive ability appropriate for individuals aged 4-90 years (Kaufman, 2004). The KBIT-2 generates a full-scale IQ (FSIQ) score, as well as Verbal and Nonverbal standard scores and age-equivalent scores. Standard scores were used for descriptive purposes and the KBIT-2 Verbal and Nonverbal age-equivalent scores were used in analyses.

The Abbreviated Battery from the SB-5 (ABIQ) is another brief assessment of cognitive ability (Roid, 2003; Roid & Barram, 2004). The SB-5 is appropriate for individuals aged 2 to over 85 years. The SB-5 ABIQ age-equivalent score was used in analyses.

The PPVT-5 is a standardized assessment of receptive language skills, specifically vocabulary knowledge of nouns, adjectives, and verbs. It is appropriate for individuals aged 2.5

years to over 90 years. Standard scores were used for descriptive purposes and age-equivalent scores for analyses.

The EVT-3 is co-normed with the PPVT-5 and is a standardized assessment of expressive language skills, specifically vocabulary and word retrieval. It is appropriate for individuals aged 2.5 years to over 90 years. Standard scores were used for descriptive purposes and age-equivalent scores for analyses. Six participants were minimally verbal and unable to complete the EVT-3 (Koegel et al., 2020).

Executive Functioning

Behavior Rating Inventory of Executive Function, second edition (BRIEF-2). The BRIEF-2 is a 63-item parent rating form of executive functioning for children 5-18 years old (Gioia et al., 2015). Items are rated on a three-point scale (i.e., never, sometimes, often) and generate age and gender standardized T-scores, with a mean of 50 and a standard deviation of 10. Higher scores represent greater difficulty with executive functioning, with a T-score of 65 serving as a clinical cut-off for potentially clinically elevated scores. An overall Global Executive Composite (GEC) is generated, as well as nine subscales, including Inhibit, Self-Monitor, Shift, Emotional Control, Initiate, Working Memory, Plan/Organize, Task-Monitor, and Organization of Materials. In previous reviews, the BRIEF-2 is recommended for use in studies that include children and youth with DS (Esbensen et al., 2019; Esbensen et al., 2017). T-scores were generated using the chronological age of the child with DS.

Behavior Rating Inventory of Executive Function, Preschool (BRIEF-P). The BRIEF-P has 63 items and is designed for children ages 2-5 years (Gioia et al., 2003). Item scoring options are the same as the BRIEF-2, and the measure also generates a GEC and T-scores. However, the BRIEF-P has only five subscale scores, including Inhibit, Shift, Emotional

Control, Working Memory, and Plan/Organize. Five T-scores for each subscale were generated using age-equivalent scores produced by each of the measures of cognition/language.

Data Analysis

The first aim of the study was to describe the appropriateness of the age-equivalent scores generated by the four measures of cognition/language to the age range of the BRIEF-P. The percentage of children with age-equivalent scores within ("age equivalent on target"), above ("age equivalent above"), or below ("age equivalent below") the chronological age range of the BRIEF-P (2:0 – 5:11 years) were reported. In addition, children with age-equivalents below 4:0 on the KBIT-2 Verbal or Nonverbal were classified as "age equivalent below" given the inability to provide an exact assessment of developmental level below this threshold. Additionally, age-equivalent scores generated from the different measures of cognition/language were compared via paired samples t-tests to identify differences and similarities between the measures and to refine which instruments were to be used to estimate age-equivalent scores for the primary analyses comparing the BRIEF-P vs BRIEF-2.

The second aim of the study was to understand the implications of using age-equivalent scores generated by measures of cognition/language to score the BRIEF-P for research purposes. For comparison across BRIEF-2 and BRIEF-P, only T-scores generated from GEC, Inhibit, Shift, Emotional Control, Working Memory and Plan/Organize were used in analyses, as the GEC and corresponding scales are shared across the two instruments. Per standard procedures, children were assigned age-equivalent scores based on their raw score on the respective measure of cognition/language. Given limitations in the age-equivalent score range for the different measures of cognition and language and the BRIEF-P itself, the following exceptions were made. First, children receiving raw scores corresponding to below the lowest age-equivalent

score possible were assigned an age equivalency score of one month less than that age in order to calculate T-scores on the BRIEF-P (e.g., an age-equivalent score of < 48 months was scored as 47 months). Second, in the case of children who received age-equivalent scores of < 2 years on the relevant cognitive measure, an age-equivalent score of 2 years was used to generate BRIEF-P T-scores (as the lower age limit of the BRIEF-P is 2 years, 0 months). Third, children receiving age-equivalent scores of 6 years or above on the measures of cognition/language were scored as 5 years 11 months on the BRIEF-P, again due to limits in the normative age range for the BRIEF-P (i.e., upper age limit for scoring is 5 years, 11 months). Correlations were run between subscale T-scores generated by the BRIEF-2 and subscale T-scores generated with age-equivalent scores from measures of cognition/language on the BRIEF-P. These correlations were run separately for all children, for children with "age equivalent on target" (i.e., their calculated age-equivalent score from the measures of cognition/language fell in the targeted age range for the BRIEF-P of 2 to 5 years) and for children with "age equivalent above" (i.e., their calculated age-equivalent score from the measures of cognition/language was higher than the upper limit of 5 years for the BRIEF-P). Children who were in the "age equivalent below" group were not evaluated as a separate group given their small distribution (n=9 on SB-5 ABIQ, n=6 on PPVT-5) but were included in the correlations for all children. Paired t-tests were run to compare the subscale T-scores generated by the BRIEF-2 and the subscale T-scores generated with age-equivalent scores from measures of cognition/language on the BRIEF-P. A threshold of acceptable correlations of $r \geq .8$ was set a priori.

The third aim of the study was to understand the implications of using age-equivalent scores generated by measures of cognition/language to score the BRIEF-P for clinical purposes. T-scores generated by the BRIEF-2 or BRIEF-P were recoded as falling above (≥ 65) or below

(< 65) the threshold for clinically elevated scores. The percentage of agreement in assigning T-scores above or below the clinical cut-off was calculated and acceptable agreement of > 90% was set a priori.

Results

Appropriateness of Age-Equivalent Score Conversion

Table 1 presents the means, standard deviations, and range for standard scores and age-equivalent scores generated by the measures of cognition and language. Six children were unable to complete the EVT-3. These children were excluded list-wise in Table 1. Age-equivalent scores generated by the SB-5 ABIQ demonstrated the highest percentage (78.9%) of children falling within the appropriate age range of the BRIEF-P, with only 11.6% receiving age-equivalent scores considered too old to be scored using the BRIEF-P. Approximately two-thirds (64.2%) of children received age-equivalent scores within the age range of the BRIEF-P when using the PPVT-5, and almost a third of children (29.5%) received age-equivalent scores considered too old to be scored using the BRIEF-P. Age-equivalent scores generated by the EVT-3 and KBIT-2 (verbal and nonverbal) had lower percentages of children falling in the age range appropriate for BRIEF-P scoring (59.6%, 34.7%, and 46.3% respectively). A higher percentage (33.7%) of children received EVT-3 age-equivalent scores considered too old to be scored using the BRIEF-P. Using KBIT-2 verbal and nonverbal age-equivalent scores, a smaller percentage (26.4% and 10.5% respectively) of children were considered too old to be scored using the BRIEF-P.

Results of paired-comparisons of the five age-equivalent scores generated by the four measures of cognition/language were evaluated to establish parsimony in subsequent analyses (see Table 2). Six children were unable to complete the EVT-3. These children were excluded list-wise in Table 2. The age-equivalent score generated by the SB-5 ABIQ was significantly

lower than those generated by all other measures of cognition/language [versus KBIT-2 Verbal, $t(94) = 10.38$, $p < .001$; versus KBIT-2 Nonverbal, $t(94) = 5.86$, $p < .001$; versus PPVT-5, $t(94) = -6.43$, $p < .001$; versus EVT-3, $t(88) = -8.54$, $p < .001$]. The age-equivalent scores generated by the PPVT-5 did not demonstrate statistically significant differences from any of the remaining measures of cognition/language [versus KBIT-2 Verbal, $t(94) = 1.83$, $p = .071$; versus KBIT-2 Nonverbal, $t(94) = -1.46$, $p = .147$; versus EVT-3, $t(88) = -0.88$, $p = .379$], which suggests that it may serve as a representative age-equivalent score across these measures of cognition/language.

In order to reduce the number of measures of cognition and language used to generate age-equivalent scores for the study's primary analyses, the percentages of participants with age-equivalent scores in the target range for the BRIEF-P (Table 1) and the results of the paired comparisons of the age-equivalent scores generated by the different measures of cognition/language (Table 2) were considered. As the SB-5 identified the most participants with age-equivalent scores in the BRIEF-P age range and the PPVT-5 age-equivalent scores did not differ significantly from the other measures of cognition/language, these two measures were used to generate age-equivalent scores for subsequent analyses that compared and contrasted the BRIEF-P and BRIEF-2. However, it is important to note that the pattern of findings throughout was comparable when assessed using age-equivalent scores generated by KBIT-2 Verbal, KBIT-2 Nonverbal and EVT-3 (data available from first author).

A visual inspection of the chronological ages of the children demonstrates that children falling in the "age equivalent above" group when scored using the SB-5 ABIQ age-equivalent scores were all chronological ages 11 and older (see Figure 1). When using the PPVT-5 age-equivalent scores, children in the "age equivalent above" group were predominantly over the

chronological age of 11 years, with one 7-year-old generating an age-equivalent score too high to score with the BRIEF-P (see Figure 2).

Comparison of Subscale T-Scores Across BRIEF-2 and BRIEF-P

Correlations Between Corresponding T-Scores on BRIEF-2 and BRIEF-P

Table 3 presents correlations between T-scores generated on subscales of the BRIEF-2 and T-scores generated on corresponding subscales of the BRIEF-P by using SB-5 ABIQ age-equivalent scores and PPVT-5 age-equivalent scores to calculate the T-scores for each form. Correlations are presented for three groupings: all children, only those children with age-equivalent scores within the age range of the BRIEF-P ("age equivalent on target"), and only those children with age-equivalent scores too high for use with the BRIEF-P ("age equivalent above"). Correlations meeting a priori thresholds were found among corresponding subscales of Inhibit, Shift, Emotional Control, Working Memory, and GEC generated by the BRIEF-2 and the BRIEF-P for all children and "age equivalent on target" children, when scored using- either the SB-5 ABIQ or PPVT-5 age-equivalent scores. In contrast, fewer correlations met the a priori threshold for "age equivalent above" children, for whom the threshold was met for only Shift, Emotional Control, and GEC when scored using SB-5 ABIQ age-equivalent scores, and for Shift, Emotional Control, Working Memory, and GEC when scored using the PPVT-5 age-equivalent scores. The correlations between the corresponding subscale of Plan/Organize generated by the BRIEF-2 and the BRIEF-P did not meet *a priori* thresholds, regardless of whether age-equivalent scores were generated from the SB-5 ABIQ or PPVT-5, and for all, "age equivalent on target," or "age equivalent above" children.

T-Test Comparing Corresponding T-Scores Between BRIEF-2 and BRIEF-P

Figure 3 illustrates the T-scores for the five subscales and GEC when scored using chronological age and the BRIEF-2 as well as mental age (calculated using the SB-5 or PPVT-5) and the BRIEF-P. Table 4 presents comparisons between BRIEF-2 subscale T-scores calculated using chronological age to the corresponding BRIEF-P subscale T-scores generated using age-equivalent scores from the SB-5 ABIQ and then the PPVT-5. T-scores for Shift were significantly higher using BRIEF-2 scoring than using BRIEF-P age-equivalent scoring for either the SB-5 ABIQ, $t(94) = -9.10$, $p < .001$, $d = 0.93$, or PPVT-5, $t(94) = -8.84$, $p < .001$, $d = 0.91$. T-scores for Emotional Control were also significantly higher using BRIEF-2 scoring than using BRIEF-P age-equivalent scoring for either the SB-5 ABIQ, $t(94) = -14.92$, $p < .001$, $d = 1.53$ or PPVT-5, $t(94) = -13.81$, $p < .001$, $d = 1.42$. T-scores for GEC were also significantly higher using BRIEF-2 scoring than using BRIEF-P age-equivalent scoring for either the SB-5 ABIQ, $t(94) = -3.99$, $p < .01$, $d = 0.34$, or PPVT-5, $t(94) = -2.36$, $p < .05$, $d = 0.24$. T-scores for Working Memory were significantly lower using BRIEF-2 scoring than using BRIEF-P age-equivalent scoring for either the SB-5 ABIQ, $t(94) = 8.11$, $p < .001$, $d = 0.83$, or PPVT-5, $t(94) = 8.37$, $p < .001$, $d = 0.86$. No statistically significant findings were identified between BRIEF-2 and BRIEF-P scoring using age-equivalent scores generated from the SB-5 ABIQ or PPVT-5 for Inhibit or Plan/Organize.

The same pattern of findings was identified for children "age equivalent on target." When assessed for "age equivalent above" children, the difference between GEC scores was not statistically significant. These analyses are not presented here for brevity and are available from the first author.

Comparison of subscale clinical cut-off agreement across BRIEF-2 and BRIEF-P

The percentage of scores falling above clincal thresholds is presented in Table 4. Table 5 presents the percentage of clinical cut-off agreement between BRIEF-2 subscale scores and corresponding subscale scores on the BRIEF-P generated using age-equivalent scores from the SB-5 ABIQ or PPVT-5. Agreement included both scores falling above or both scores falling below clinical cut-offs. Agreement was found to exceed the a priori 90% threshold for the Inhibit, Emotional Control, and GEC subscales when assessed using either the SB-5 ABIQ or PPVT-5 age-equivalent scoring. This pattern of agreement was present when evaluated with all children, and both the "age equivalent on target," and "age equivalent above" subgroups. Agreement also exceeded the a priori threshold of 90% agreement for the Shift, Working Memory, and Plan/Organize subscales for "age equivalent above" children when using age-equivalent scores generated using the SB-5 ABIQ, but not the PPVT-5.

Discussion

The current study evaluated the appropriateness of scoring the BRIEF-P using age-equivalent scores generated from different measures of cognition/language among school-age children with DS. Analyses also examined how T-scores generated by the BRIEF-P using these age-equivalent scores compared to T-scores generated by the BRIEF-2 based on the child's chronological age. As age-equivalent scores for the PPVT-5 were similar to the different measures of cognition/language (KBIT-2, EVT-3), and age-equivalents for the SB-5 ABIQ were distinct from the other measures of cognition/language, subsequent analyses only examined T-scores calculated using age-equivalent scores generated by the SB-5 ABIQ and PPVT-5 for parsimony. Findings demonstrated that when using chronological age for BRIEF-2 scoring or age-equivalent scores for BRIEF-P scoring, T-scores exhibited areas of consistency and areas of

disagreement across these scoring methods. These areas of consistency and disagreement have implications for research and clinical practice.

T-scores demonstrated consistency in intercorrelations when generated using chronological age on the BRIEF-2 or age-equivalent scores on the BRIEF-P. T-scores on the subscales of Inhibit, Shift, Emotional Control, Working Memory, and the GEC all demonstrated strong intercorrelations with their respective subscale across the BRIEF-2 and BRIEF-P scoring, meeting a priori thresholds. However, if age-equivalent scores were considered "too high" for scoring on the BRIEF-P, these intercorrelations weakened, and strong inter-correlations were only found for T-scores on the subscales of Shift, Emotional Control, and the GEC. T-scores on the Plan/Organize subscale did not demonstrate comparably strong intercorrelations for any children.

Less agreement was identified among T-scores generated using chronological age on the BRIEF-2 and age-equivalent scores on the BRIEF-P when mean T-scores on each instrument were compared. When comparing mean BRIEF-2 T-scores with mean BRIEF-P T-scores generated using age-equivalent scores from SB-5 ABIQ or PPVT-5, only the Inhibit and Plan/Organize subscales demonstrated comparable scores. The mean BRIEF-P T-scores generated by the age-equivalent scores from the SB-5 ABIQ and PPVT-5 were comparable to each other, yet both were consistently lower than mean BRIEF-2 T-scores on the subscales of Shift, Emotional Control, and GEC, and consistently higher than mean BRIEF-2 T-scores on the Working Memory subscale. This pattern of findings suggests comparable skills in school-age children with DS to preschoolers without disabilities (reflected by scores on the BRIEF-P) in self-regulatory skills related to flexibility and emotional control. Yet when compared to school-age children without disabilities (reflected by scores on the BRIEF-2), the disparity in

expectations is greater and children with DS are scored as having greater difficulties with shifting and emotional control.

The implications for research are that the BRIEF-P can be scored for school-age children with DS using age-equivalent scores generated from measures of cognition/language, with confidence in the Inhibit subscale T-scores. It is recommended that when this alternate scoring method is used, the Plan/Organize T-scores *not* be considered valid for statistical analyses given weaker intercorrelations between BRIEF measures. The T-scores for subscales of Shift, Emotional Control, Working Memory, and GEC can be used for research analyses, with recognition that scores may be consistently higher or lower than if used age-appropriately (see Figure 3). These findings provide guidance for which measures of executive functioning to use for different age ranges of children with DS, and guidance of which measures to use for future research that spans ages that extend across the age range of the BRIEF-P or BRIEF-2. These findings corroborate others that have suggested the potential for using the BRIEF-P with age-equivalent scores among adolescents with DS (d'Ardhuy et al., 2015).

These findings replicate other examinations of the BRIEF measures. The pattern of strengths and weaknesses in BRIEF-2 subscale T-scores among *school-age children* with DS demonstrated in Figure 3 was consistent with that reported in the literature also when using the BRIEF-2 with *school-age children* with DS (Onnivello et al., 2021). Further, the pattern of strengths and weaknesses in BRIEF-P subscale T-scores generated from age-equivalent scores using measures of cognition/language among *school-age children* with DS was consistent with that reported in the literature when using the BRIEF-P age-appropriately with *preschoolers* with DS (Onnivello et al., 2021). These consistencies with profiles reported in the literature replicate

prior findings, yet longitudinal studies are warranted to better understand developmental changes in subdomains of executive functioning in children with DS.

When examining cut-off scores, we identified some limitations for use of BRIEF-P with age-equivalent scores generated from measures of cognition/language among school-age children with DS. Alternate scoring on the BRIEF-P demonstrated sound agreement for decisions using clinical cut-off scores of 65 with the subscales of Inhibit, Emotional Control, and GEC, but not for the subscales of Shift, Working Memory, or Plan/Organize. The implications for clinic settings are to use more caution if using the BRIEF-P with school-age children, as agreement is below 90% and can fall as low as 73%. Thus, if a clinician is observing a child multiple times from preschool to school-age and wishes to use the same measures across assessments, the BRIEF-P could continue to be used, albeit with reduced confidence in interpretation of subscales of Shift, Working Memory, and Plan/Organize. However, if longitudinal information is not being assessed clinically, then clinicians are advised to use the age-appropriate version of the BRIEF.

Several study limitations should be noted. First, for parsimony, only findings from BRIEF-P generated from age-equivalent scores on the SB-5 ABIQ and PPVT-5 were presented. Although comparable patterns of findings were found with the age-equivalent scores on the EVT-3, KBIT-2 Verbal and KBIT-2 Nonverbal, restrictions on the age-equivalents on the KBIT-2 warrant caution. As the lowest age-equivalent score generated from the KBIT-2 is < 4 years 0 months, this measure is limited in its ability to create age-equivalent scores across the age range of the BRIEF-P, and it not recommended for use in alternate BRIEF-P scoring. Second, the assessment order of measures of cognition and language were not counterbalanced or randomized. And third, using age-equivalent scores is a limitation of the current study. Age-

equivalent scores have limited reliability and validity as they are based on median age scores for corresponding raw scores (Bracken, 1988). Nonetheless, the consistency of the current findings across age-equivalent scores from the SB-5 ABIQ and PPVT-5, together with cross-sectional findings on the BRIEF-P and BRIEF-2 (Onnivello et al., 2021), supports our current recommendations.

Our study corroborates prior findings regarding the T-score subscale profile of the BRIEF-P and BRIEF-2, regardless of whether the BRIEF-2 is used age-appropriately or whether the BRIEF-P is scored using age-equivalent scores generated from measures of cognition/language. Our study also extends the literature by providing recommendations for research and clinical practice with BRIEF-P generated from age-equivalent scores. Overall, we recommend that future research and clinical practice use (1) the BRIEF-P for children with DS ages 2-5 years old, (2) the BRIEF-2 with chronological-age scoring for research and clinic, or the BRIEF-P with age-equivalent scoring (with some caveats) for research for children with DS ages 5-10 years old, and (3) the BRIEF-2 for children with DS ages 11 and older. Caveats for using the BRIEF-P with age-equivalents include not using the Plan/Organize T-scores and attending to previously described recommendations related to interpretation of specific subscale scores.

References

Baddeley, A., & Jarrold, C. (2007). Working memory and Down syndrome. *Journal of Intellectual Disability Research*, *51*(12), 925-931.

Bracken, B. A. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology*, *26*(2), 155-166.

Camp, J. S., Karmiloff-Smith, A., Thomas, M. S., & Farran, E. K. (2016). Cross-syndrome comparison of real-world executive functioning and problem solving using a new problem-solving questionnaire. *Research in developmental disabilities*, *59*, 80-92.

d'Ardhuy, X. L., Edgin, J. O., Bouis, C., de Sola, S., Goeldner, C., Kishnani, P., Nöeldeke, J., Rice, S., Sacco, S., Squassante, L., Spiridigliozzi, G. A., Visootsak, J., Heller, J. H., & Khwaja, O. (2015). Assessment of cognitive scales to examine memory, executive function and language in individuals with Down syndrome: implications of a 6-month observational study. *Frontiers in Behavioral Neuroscience*, *9*, 300.

Daunhauer, L. A., Fidler, D. J., Hahn, L., Will, E., Lee, N. R., & Hepburn, S. (2014). Profiles of everyday executive functioning in young children with Down syndrome. *American journal on intellectual and developmental disabilities*, *119*(4), 303-318.

Daunhauer, L. A., Gerlach-McDonald, B., Will, E., & Fidler, D. (2017). Performance and ratings based measures of executive function in school-aged children with Down syndrome. *Developmental Neuropsychology*, *42*(6), 351-368.

Diamond, A. (2013). Executive functions. *Annual review of psychology*, *64*, 135-168.

DiStefano, C., Sadhwani, A., & Wheeler, A. C. (2020). Comprehensive assessment of individuals with significant levels of intellectual disability: challenges, strategies, and future directions. *American journal on intellectual and developmental disabilities*, *125*(6), 434-448.

Dunn, D. M. (2019). *Peabody Picture Vocabulary Test, Fifth Edition*. Pearson.

Edgin, J. O., Anand, P., Rosser, T., Pierpont, E. I., Figuero, C., Hamilton, D., Huddleston, L., Mason, G. M., Spanò, G., Toole, L., Nguyen-Driver, M., Capone, G., Abbeduto, L., Maslen, C., Reeves, R. H., & Sherman, S. L. (2017). The Arizona Cognitive Test Battery for Down Syndrome: Test-retest reliability and practice effects. *American Journal of Intellectual and Developmental Disabilities*, *122*(3), 215-234.

Edgin, J. O., Mason, G. M., Allman, M. J., Capone, G. T., DeLeon, I., Maslen, C., Reeves, R. H., Sherman, S. L., & Nadel, L. (2010). Development and validation of the Arizona Cognitive Test Battery for Down syndrome. *Journal of neurodevelopmental disorders*, *2*, 149-164.

Esbensen, A. J., Hoffman, E., Shaffer, R., Chen, E., Patel, L., & Jacola, L. M. (2019). Reliability of informant report measure of executive functioning in children with Down syndrome. *American Journal of Intellectual and Developmental Disabilities*, *124*, 220-233.

Esbensen, A. J., Hoffman, E. K., Shaffer, R. C., Patel, L. R., & Jacola, L. M. (2021). Relationship Between Parent and Teacher Reported Executive Functioning and Maladaptive Behaviors in Children With Down Syndrome. *American journal on intellectual and developmental disabilities*, *126*(4), 307-323.

Esbensen, A. J., Hooper, S. R., Fidler, D., Hartley, S. L., Edgin, J. O., d'Ardhuy, X. L., Capone, G., Conners, F. A., Mervis, C. B., Abbeduto, L., Rafii, M., Krinsky-McHale, S. J., & Urv, T. (2017). Outcome measures for clinical trials in Down syndrome. *American journal on intellectual and developmental disabilities*, *122*(3), 247-281.

Gioia, G. A., Espy, K. A., & Isquith, P. K. (2003). *BRIEF-P: Behavior Rating Inventory of Executive Function--Preschool Version: Professional Manual*. Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2015). *Behavior Rating Inventory of Executive Function 2nd Edition (BRIEF2): Professional Manual*. Psychological Assessment Resources, Incorporated.

Kaufman, A. (2004). KBIT-2: Kaufman Brief Intelligence Test. Second Edition. In. Upper Saddle River, NJ: Pearson.

Kittler, P., Krinsky- McHale, S., & Devenny, D. (2006). Verbal intrusions precede memory decline in adults with Down syndrome. *Journal of Intellectual Disability Research*, *50*(1), 1-10.

Koegel, L. K., Bryan, K. M., Su, P. L., Vaidya, M., & Camarata, S. (2020). Definitions of nonverbal and minimally verbal in research for autism: A systematic review of the literature. *Journal of Autism and Developmental Disorders*, *50*, 2957-2972.

Lanfranchi, S., Cornoldi, C., Vianello, R., & Conners, F. (2004). Verbal and visuospatial working memory deficits in children with Down syndrome. *American Journal on Mental Retardation*, *109*(6), 456-466.

Lanfranchi, S., Jerman, O., Dal Pont, E., Alberti, A., & Vianello, R. (2010). Executive function in adolescents with Down Syndrome. *Journal of Intellectual Disability Research*, *54*, 308-319.

Lanfranchi, S., Jerman, O., & Vianello, R. (2009). Working memory and cognitive skills in individuals with Down syndrome. *Child Neuropsychology*, *15*(4), 397-416.

Lee, N. R., Anand, P., Will, E., Adeyemi, E. I., Clasen, L. S., Blumenthal, J. D., Giedd, J. N., Daunhauer, L. A., Fidler, D. J., & Edgin, J. O. (2015). Everyday executive functions in Down syndrome from early childhood to young adulthood: evidence for both unique and shared characteristics compared to youth with sex chromosome trisomy (XXX and XXY). *Frontiers in Behavioral Neuroscience*, *9*, 264.

Lee, N. R., Fidler, D. J., Blakeley-Smith, A., Daunhauer, L., Robinson, C., & Hepburn, S. L. (2011). Caregiver report of executive functioning in a population-based sample of young children with Down syndrome. *American journal on intellectual and developmental disabilities*, *116*(4), 290-304.

Loveall, S., Conners, F., Tungate, A., Hahn, L., & Osso, T. (2017). A cross- sectional analysis of executive function in Down syndrome from 2 to 35 years. *Journal of Intellectual Disability Research*, *61*(9), 877-887.

Memisevic, H., & Sinanovic, O. (2014). Executive function in children with intellectual disability–the effects of sex, level and aetiology of intellectual disability. *Journal of Intellectual Disability Research*, *58*(9), 830-837.

Onnivello, S., Colaianni, S., Pulina, F., Locatelli, C., Marcolin, C., Ramacieri, G., Antonaros, F., Vione, B., Piovesan, A., & Lanfranchi, S. (2021). Executive functions and adaptive behaviour in individuals with Down syndrome. *Journal of Intellectual Disability Research*.

Pennington, B. F., Moon, J., Edgin, J. O., Stedron, J., & Nadel, L. (2003). The neuropsychology of Down syndrome: evidence for hippocampal dysfunction. *Child development*, *74*(1), 75-93.

Pritchard, A., Kalback, S., McCurdy, M., & Capone, G. T. (2015). Executive functions among youth with Down Syndrome and co- existing neurobehavioural disorders. *Journal of Intellectual Disability Research*, *59*(12), 1130-1141.

Roid, G. H. (2003). *Stanford-binet intelligence scales (SB5)*. Riverside.

Roid, G. H., & Barram, R. A. (2004). *Essentials of Stanford-Binet intelligence scales (SB5) assessment* (Vol. 39). John Wiley & Sons.

Rowe, J., Lavender, A., & Turk, V. (2006). Cognitive executive function in Down's syndrome. *British Journal of Clinical Psychology*, *45*(1), 5-17.

Schworer, E., Voth, K., Hoffman, E., & Esbensen, A. (2022). Short-term memory outcome measures: Psychometric evaluation and performance in youth with Down syndrome. *Research in developmental disabilities*, *120*, 104147.

Schworer, E. K., Fidler, D., Lunkenheimer, E., & Daunhauer, L. (2019). Parenting behaviour and executive function in children with Down syndrome. *Journal of Intellectual Disability Research*, *63*(4), 298-312.

Tungate, A. S., & Conners, F. A. (2021). Executive function in Down syndrome: A meta-analysis. *Research in developmental disabilities*, *108*, 103802.

Vicari, S., Carlesimo, A., & Caltagirone, C. (1995). Short-term memory in persons with intellectual disabilities and Down's syndrome. *Journal of Intellectual Disability Research*, *39*, 307-321.

Will, E., Fidler, D., Daunhauer, L., & Gerlach- McDonald, B. (2017). Executive function and academic achievement in primary–grade students with Down syndrome. *Journal of Intellectual Disability Research*, *61*(2), 181-195.

Will, E. A., Schworer, E. K., & Esbensen, A. J. (2021). The role of distinct executive functions on adaptive behavior in children and adolescents with Down syndrome. *Child Neuropsychology*, 1-19.

Williams, K. T. (2018). *Expressive vocabulary test, third edition* Pearson.

Table 1.

Standard scores and age-equivalent scores generated by measures of cognition/language.

	Standard Score		Age-Equivalent Score (months)		Number and percent with age-equivalent score too low for BRIEF-P	Number and percent with age-equivalent score appropriate for BRIEF-P	Number and percent with age-equivalent score too high for BRIEF-P
	Mean (SD)	Range	Mean (SD)	Range			
KBIT-2 Verbal	49.56 (11.14)	40-82	61.94 (17.38)	47-114	37; 38.9%	33; 34.7%	25; 26.3%
KBIT-2 Nonverbal	49.25 (11.26)	40-90	56.68 (13.52)	47-124	41; 43.2%	44; 46.3%	10; 10.5%
SB-5 ABIQ	49.36 (5.59)	47-76	49.50 (16.72)	24-100	9; 9.5%	74; 78.9%	11; 11.6%
PPVT-5	50.93 (13.86)	20-87	59.11 (22.95)	24-130	6; 6.3%	61; 64.2%	28; 29.5%
EVT-3	62.10 (11.99)	20-85	62.12 (18.38)	29-107	6; 6.7%	53; 59.6%	30; 33.7%

Note. BRIEF-P: Behavior Rating Inventory of Executive Function, preschool version; EVT-3: Expressive Vocabulary Test, third edition; KBIT-2: Kaufman Brief Intelligence Test – second edition; PPVT-5: Peabody Picture Vocabulary Test, fifth edition; SB-5-ABIQ: Stanford-Binet fifth edition Abbreviated Battery IQ.

Table 2.

Paired comparison between age-equivalent scores generated by different measures of cognition/language.

	KBIT-2 Verbal	KBIT-2 Nonverbal	SB-5 ABIQ	PPVT-5
KBIT-2 Verbal				
KBIT-2 Nonverbal	$t(94) = 3.84$***			
SB-5 ABIQ	$t(94) = 10.38$***	$t(94) = 5.86$***		
PPVT-5	$t(94) = 1.83$	$t(94) = -1.46$	$t(94) = -6.43$***	
EVT-3	$t(88) = 0.36$	$t(88) = -3.15$**	$t(88) = -8.54$***	$t(88) = -0.88$

***$p < .001$, ** $p < .01$

Note. EVT-3: Expressive Vocabulary Test, third edition; KBIT-2: Kaufman Brief Intelligence Test – second edition; PPVT-5: Peabody Picture Vocabulary Test, fifth edition; SB-5-ABIQ: Stanford-Binet, fifth edition, Abbreviated Battery IQ.

Table 3.

*Correlations between corresponding subscales on BRIEF-2 and BRIEF-P (generated using SB-5 ABIQ and PPVT-5 age-equivalent scores) for all children, children "*age equivalent *on target", and children "*age equivalent above*" for using BRIEF-P.*

BRIEF-2	BRIEF-P	
	Corresponding subscale using SB-5 ABIQ age-equivalent scores	Corresponding subscale using PPVT-5 age-equivalent scores
All children	n=95	n=95
Inhibit	**.84*** **	**.84*** **
Shift	**.86*** **	**.85*** **
Emotional Control	**.93*** **	**.93*** **
Working Memory	**.90*** **	**.89*** **
Plan/Organize	.53***	.50***
GEC	**.93*** **	**.92*** **
"Age equivalent on target"	n=75	n=61
Inhibit	**.86*** **	**.87*** **
Shift	**.86*** **	**.82*** **
Emotional Control	**.91*** **	**.93*** **
Working Memory	**.91*** **	**.89*** **
Plan/Organize	.57***	.60***
GEC	**.93*** **	**.92*** **
"Age equivalent above"	n=11	n=28
Inhibit	.61*	.72***
Shift	**.93*** **	**.95*** **
Emotional Control	**.98*** **	.94***
Working Memory	.75**	.85***
Plan/Organize	.62*	.40*
GEC	**.93*** **	**.94*** **

***$p < .001$, ** $p < .01$, * $p < .05$

Note. Bolding indicates meeting a prior criterion: $r \geq .8$

Note. GEC: Global Executive Composite; PPVT-5: Peabody Picture Vocabulary Test, fifth edition; SB-5-ABIQ: Stanford-Binet, fifth edition, Abbreviated Battery IQ.

Table 4.

Means T-scores generated and percentage above clinical thresholds for all children on BRIEF-P (using age-equivalent score generated from SB-5 ABIQ and PPVT) and BRIEF-2 (using chronological age) (n=95).

	BRIEF-P score based on age-equivalent score generated from:				BRIEF-2 score based on chronological age	
	SB-5 ABIQ		PPVT-5			
	M (SD)	% above clinical threshold	M (SD)	% above clinical threshold	M (SD)	% above clinical threshold
Inhibit	55.2 (10.7)	22.1%	55.3 (10.7)	23.2%	55.6 (10.2)	20.0%
Shift	54.8 (10.7)[a]	20.0%	55.0 (10.7)[b]	20.0%	60.5 (11.6)[a,b]	38.9%
Emotional Control	45.2 (8.9)[a]	4.2%	45.6 (9.4)[b]	4.2%	50.8 (9.7)[a,b]	11.6%
Working Memory	66.3 (11.5)[a]	46.3%	66.5 (11.4)[b]	46.3%	61.8 (9.0)[a,b]	35.8%
Plan/Organize	56.3 (11.8)	22.1%	56.6 (11.7)	24.2%	56.8 (9.7)	15.8%
GEC	57.7 (11.3)[a]	30.5%	58.1 (11.5)[b]	31.6%	59.2 (8.9)[a,b]	29.5%

[a] Significantly different T-scores between BRIEF-P score generated from SB-5 ABIQ age-equivalent score and BRIEF-2 using chronological age.

[b] Significantly different T-scores between BRIEF-P score generated from PPVT-5 age-equivalent score and BRIEF-2 using chronological age.

Note. BRIEF-P: Behavior Rating Inventory of Executive Function, preschool version; BRIEF-2: Behavior Rating Inventory of Executive Function, second edition; GEC: Global Executive Composite; PPVT-5: Peabody Picture Vocabulary Test, fifth edition; SB-5-ABIQ: Stanford-Binet, fifth edition, Abbreviated Battery IQ.

Table 5.

Percentage of agreement across BRIEF-2 and BRIEF-P scoring with age-equivalent scores on scores falling above or below clinical threshold.

	SB-5 ABIQ			PPVT-5		
	All kids (n=95)	"Age equivalent on target" (n=75)	"Age equivalent above" (n=11)	All kids (n=95)	"Age equivalent on target" (n=61)	"Age equivalent above" (n=28)
Inhibit	**94.7%**	**96.0%**	**100%**	**93.6%**	**91.8%**	**100%**
Shift	76.8%	78.7%	**90.9%**	76.8%	73.8%	85.7%
Emotional Control	**92.6%**	**96.0%**	**90.9%**	**92.6%**	**91.8%**	**96.4%**
Working Memory	85.3%	82.7%	**100%**	85.3%	85.2%	85.7%
Plan/Organize	81.1%	82.7%	**90.9%**	76.8%	77.0%	78.6%
GEC	**93.7%**	**94.7%**	**90.9%**	**91.6%**	**90.2%**	**96.4%**

Note. Bolding indicates meeting a prior criteria: agreement >90%

Note. GEC: Global Executive Composite; PPVT-5: Peabody Picture Vocabulary Test, fifth edition; SB-5-ABIQ: Stanford-Binet, fifth edition, Abbreviated Battery IQ.

Figure 1.

Visual inspection of children "age equivalent on target" (darker shading) and "age equivalent above" (lighter shading) for scoring on the BRIEF-P when age-equivalents are generated from SB-5 ABIQ.

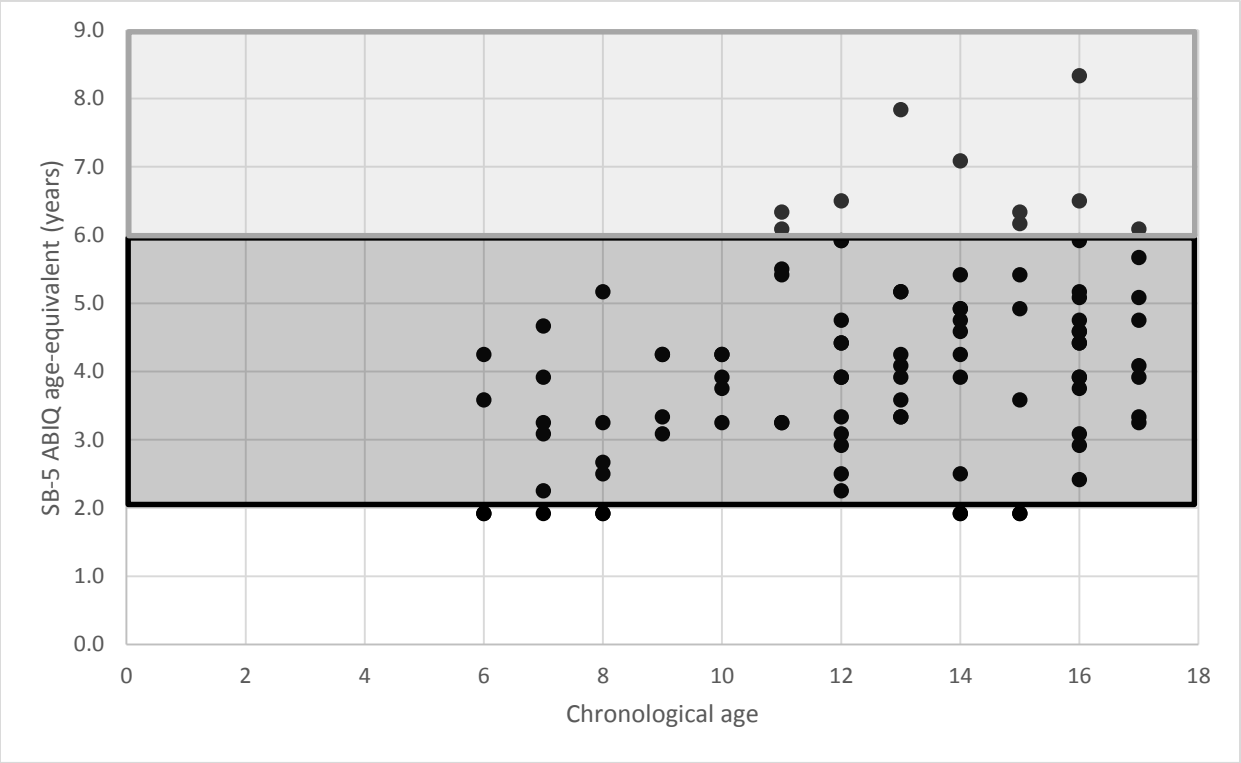


Figure 2.

Visual inspection of children "age equivalent on target" (darker shading) and "age equivalent above" (lighter shading) for scoring on the BRIEF-P when age-equivalents are generated from PPVT-5.

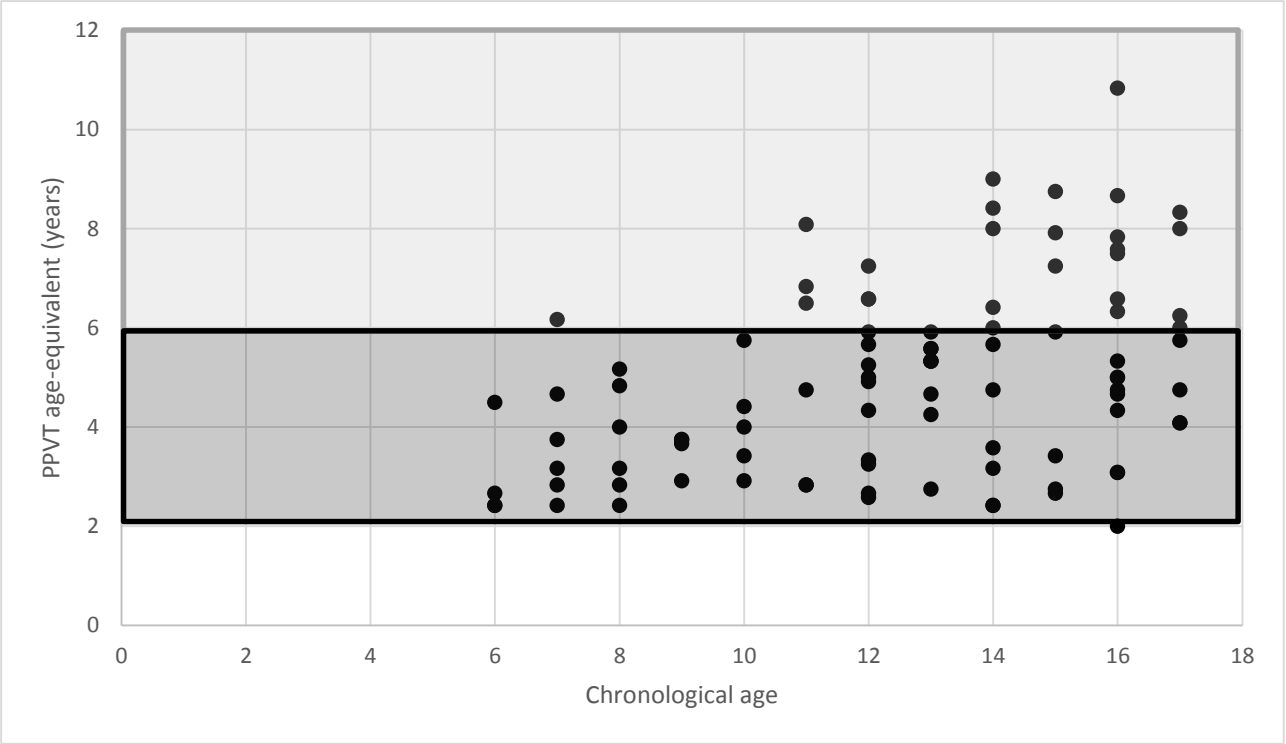
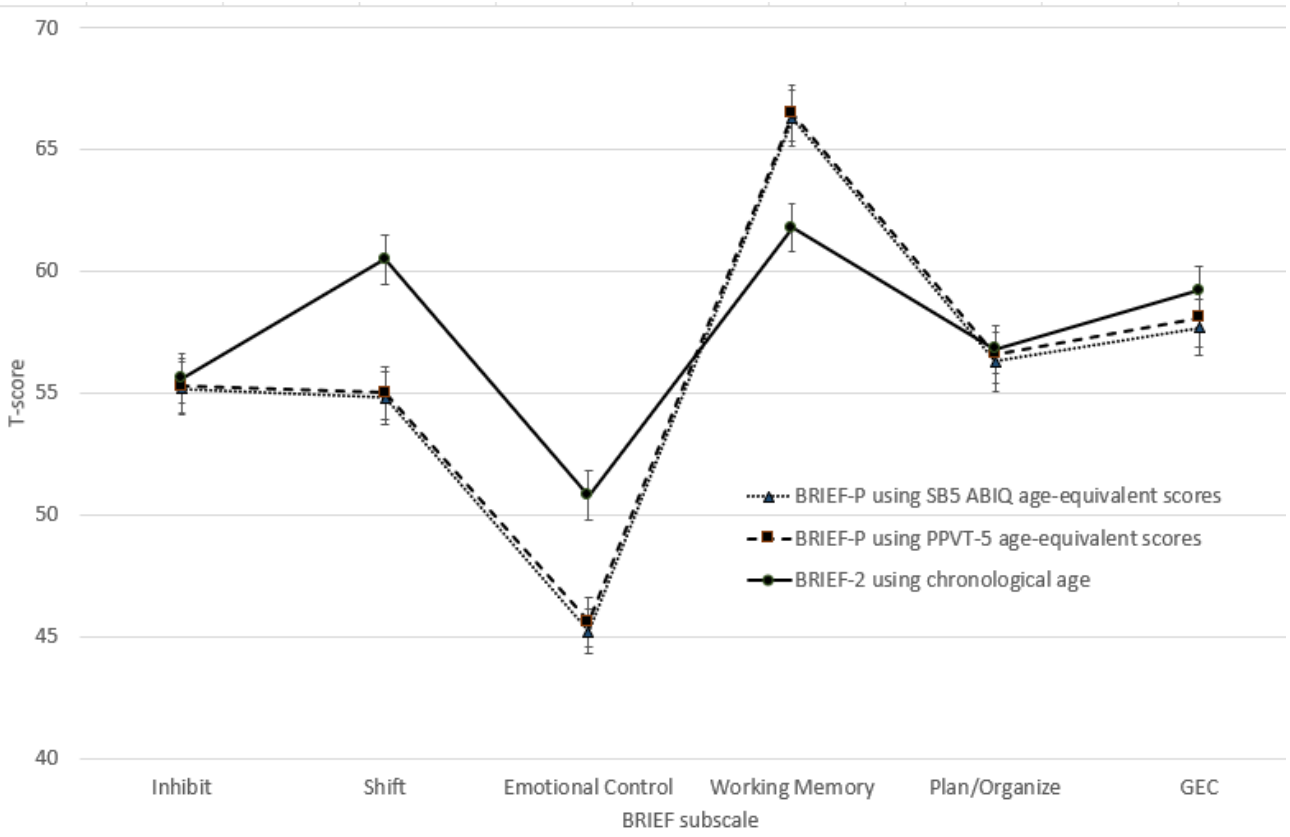


Figure 3.

T-scores and standard error bars by subscale for BRIEF-2 scored using chronological age, and BRIEF-P scored using age-equivalent scores generated from SB-5 ABIQ or PPVT-5.



April 18, 2023

Dear Dr. Finestack,

We are grateful for your attention to detail. We have incorporated changes to reflect suggestions related to Koegel et al. and highlighted the changes within the manuscript in green font. Here is a point-by-point response to the reviewer's comments and concerns.

EDITOR
The authors note that six participants were nonverbal and therefore did not complete the EVT-3. Specify what is meant by nonverbal, as this term can have different meanings coming from different researchers. See the following article for a discussion around this issue in autism.

Koegel LK, Bryan KM, Su PL, Vaidya M, Camarata S. Definitions of Nonverbal and Minimally Verbal in Research for Autism: A Systematic Review of the Literature. J Autism Dev Disord. 2020 Aug;50(8):2957-2972. doi: 10.1007/s10803-020-04402-w. PMID: 32056115; PMCID: PMC7377965.

With your response as follows:
As children were older than 6 years and did not have the language skills to complete the EVT-3, in the absence of other language testing, they meet the classification for nonverbal. We have reviewed our files and corroborated that these children had very few expressive words. We have added the citation to Koegel to ensure readers understand our nomenclature (page 10).

According to Koegel et al., children would be classified as nonverbal if they are over 18 months of age and demonstrate no consistent verbal expressive words (intelligible or approximations) during standardized tests, across settings during observations, and according to parent report." While children would be classified as minimally verbal if they are "using some words, but significantly fewer than expected levels relative to age." The latter seems to describe your participants more so, but they could also be characterized as limited verbal. This should be clarified. Also please note the Koegel et al.'s reporting recommendations, "credible estimates of word counts should be included for MV individuals and credible procedures identifying individuals as NV should also be included in future studies."

Once this concern is addressed, I will be able to recommend your manuscript for publication.

- We have reclassified these children as minimally verbal on page 10. Unfortunately, in our study we do not have word count estimates but are grateful for this recommendation and will add this characteristic to subsequent studies.
    - "Standard scores were used for descriptive purposes and age-equivalent scores for analyses. Six participants were minimally verbal and unable to complete the EVT-3 (Koegel et al., 2020)."