Original Article

A comparison of WISC-IV and SB-5 intelligence scores in adolescents with autism spectrum disorder

Autism
1–10

DOI: 10.1177/1362361314554920
aut.sagepub.com

Katherine T Baum[1], Paula K Shear[2], Steven R Howe[2] and Somer L Bishop[3]

Abstract

In autism spectrum disorders, results of cognitive testing inform clinical care, theories of neurodevelopment, and research design. The Wechsler Intelligence Scale for Children and the Stanford–Binet are commonly used in autism spectrum disorder evaluations and scores from these tests have been shown to be highly correlated in typically developing populations. However, they have not been compared in individuals with autism spectrum disorder, whose core symptoms can make testing challenging, potentially compromising test reliability. We used a within-subjects research design to evaluate the convergent validity between the Wechsler Intelligence Scale for Children, 4th ed., and Stanford–Binet, 5th ed., in 40 youth (ages 10–16 years) with autism spectrum disorder. Corresponding intelligence scores were highly correlated ($r = 0.78$ to 0.88), but full-scale intelligence quotient (IQ) scores ($t(38) = -2.27$, $p = 0.03$, $d = -0.16$) and verbal IQ scores ($t(36) = 2.23$, $p = 0.03$; $d = 0.19$) differed between the two tests. Most participants obtained higher full-scale IQ scores on the Stanford–Binet, 5th ed., compared to Wechsler Intelligence Scale for Children, 4th ed., with 14% scoring more than one standard deviation higher. In contrast, verbal indices were higher on the Wechsler Intelligence Scale for Children, 4th ed., Verbal–nonverbal discrepancy classifications were only consistent for 60% of the sample. Comparisons of IQ test scores in autism spectrum disorder and other special groups are important, as it cannot necessarily be assumed that convergent validity findings in typically developing children and adolescents hold true across all pediatric populations.

Keywords

Measurement, IQ, convergent validity, cognitive, autism spectrum disorder

Measurement of intelligence in individuals with autism spectrum disorder (ASD) is integral to both clinical practice and research. Diagnostic decision-making, treatment planning, research design, and neurodevelopmental theory in ASD all rely on intelligence test scores. Despite widespread use of intelligence tests in this population, there are still many unanswered questions about how to reliably assess this construct in ASD. Children with ASD can struggle in testing situations and core symptoms of the disorder may interfere with test reliability. The social, communication, and behavioral difficulties inherent to ASD may differentially affect performance across measures. The present methodological study focused on how two widely used intelligence measures compare in characterizing the intelligence of adolescents with ASD.

Assessment of individuals with ASD often includes, as a core component, measurement of an individual's overall intellectual functioning. One major purpose of this intelligence assessment is to determine whether individuals meet the ASD diagnostic criterion that requires social communication impairments to be below what is expected given the individual's cognitive/developmental level (American Psychiatric Association, 2013). Intelligence test scores are also administered to assess possible intellectual disability (ID), which is present in an estimated 55%–70% of individuals with ASD (Fombonne, 2005).

[1]Cincinnati Children's Hospital Medical Center, USA
[2]University of Cincinnati, USA
[3]Weill Cornell Medical College, USA

Corresponding author:
Katherine T Baum, Division of Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, MLC 3015, Cincinnati, OH 45229, USA.
Email: katherine.baum@cchmc.org

Intelligence (as measured by an intelligence quotient or IQ) has strong prognostic implications in that those with ASD who have an IQ above 70 are more likely to have positive functional outcomes in adulthood, including employment, higher education, rewarding social relationships, and independent living (Howlin et al., 2004). Furthermore, an IQ cut-off score of 70 impacts service eligibility. Individuals on the autism spectrum whose IQ scores are at or below 70, indicating ID, may be afforded additional services within school or community settings, and there is evidence that decisions about educational placement and services tend to be based more strongly on an individual's cognitive abilities than on other areas of functioning, such as degree of social deficit (White et al., 2007).

From a more theoretical perspective, intelligence data have strongly impacted hypotheses about developmental cognition in ASD. Specific cognitive profiles, defined by the discrepancy between verbal IQ (VIQ) and nonverbal IQ (NVIQ) scores, are thought by some researchers to represent specific subtypes of autism (Joseph et al., 2002; Munson et al., 2008). Specifically, it has been suggested that a cognitive profile characterized by discrepant verbal and nonverbal abilities (with either nonverbal or verbal abilities being greater) using both the Wechsler scales (Wechsler, 1991, 1999, 2003) and Differential Abilities Scales (DAS) (Elliott, 1990) are associated with increased social impairment, possibly reflecting a neurobehavioral marker for an etiologically distinct subtype of autism (Black et al., 2009; Joseph et al., 2002). Consistent with this hypothesis, children with nonverbal abilities that are significantly higher than verbal abilities are shown to have a greater head circumference than those with greater verbal abilities or no significant cognitive differences (Deutsch and Joseph, 2003), and specific chromosome chains may be related to IQ-discrepancy scores (Chapman et al., 2011). These conclusions that draw upon cognitive profiles (i.e. verbal–nonverbal discrepancies) to postulate the existence of distinct subtypes of ASD have major implications for our understanding of the disorder.

IQ scores also feature prominently in research design in this population, for example, when selecting samples (e.g. restricting studies to "high-functioning" individuals) or matching comparison groups. However, there is little consistency with regard to test selection. Large-scale studies involving individuals across a wide age or developmental range are unable to use the same test for all participants due to floor effects and restrictions on test age or developmental functioning. Furthermore, studies that involve multiple sites may not use the same intelligence measure at all sites. Thus, research studies commonly draw on scores from multiple measures of intelligence, with individuals in the sample receiving different IQ tests based on convenience, their age, and/or their cognitive level (Chapman et al., 2011; Charman et al., 2011; Huber, 2008; Mayes and Calhoun, 2003). Results of these studies assume agreement between different IQ test scores without clear evidence of convergent validity, which is an identified area of concern (see also Bishop et al., 2011).

Beyond test selection, the indices selected to estimate intellectual ability (e.g. full-scale intelligence quotient (FSIQ), VIQ, NVIQ) also vary across studies. Some studies match participants using NVIQ (e.g. Howlin, 2003), while others select their sample based on VIQ (e.g. Klin et al., 2007) or both verbal and FSIQ (e.g. Goldstein et al., 2002). NVIQ scores tend to be the most stable over time in individuals with ASD (Begovac et al., 2009; Howlin et al., 2004), lending support to use of NVIQ in research. In the general population however, NVIQ scores on different tests are more weakly correlated than full-scale or VIQ scores (Roid, 2003). It is understandable that researchers elect to use different IQ scores based on their research question or on the scores most commonly used in previous studies on similar topics and it is unlikely that a single type of score will be useful for all purposes. However, it is not yet known whether one type of score exhibits higher convergent validity across tests in individuals with ASD.

There has been some debate about which of the many tools available best measure intellectual ability in children with ASD. In verbal individuals who function above the moderately impaired range of ID, the Wechsler Intelligence Scales are the most commonly administered, followed by the Stanford–Binet (SB) (Roid, 2003; Thorndike et al., 1986). The Wechsler Intelligence Scale for Children (WISC) is the most widely used to match IQ (Mottron, 2004) and is reportedly preferable because it measures multiple constructs, and its results readily translate into intervention objectives (Klin et al., 2005). Despite this support for the widespread use of the WISC in general pediatric evaluations, until recently the Autism Treatment Network (ATN), which is the nation's first network of hospitals and physicians dedicated to developing a model of comprehensive medical care for individuals with ASD, required the SB, 5th ed. (SB-5) (Roid, 2003) for assessment of cognitive functioning (Autism Speaks, 2010). The ATN now accepts several intelligence measures, including both SB and WISC, but for many years the SB-5 was the standard for children assessed at participating institutions.

The WISC and SB have been used to estimate intelligence for individuals with ASD within a given study (Huber, 2008; Mayes and Calhoun, 2003), but no studies have directly compared performance on the two measures in the same group of participants. Several reports, including the original test validation studies, have compared scores on the SB and WISC in other pediatric populations. The technical manual for the most recent edition of the SB (SB-5) reports strong correlations with WISC, 3rd ed. (WISC-III) (Wechsler, 1991) index scores in typically developing children. Specifically, the FSIQ scores on the SB-5 and WISC-III ($r = 0.84$) and the VIQ scores ($r = 0.85$)

were more highly correlated than the nonverbal scores (SB-5 NVIQ and WISC-III Performance IQ; $r=0.66$). To our knowledge, these comparisons have not been conducted with the most recent children's version of the Wechsler scale, the WISC-IV, or in children with ASD.

In several pediatric populations, including those with mild ID and those with academic difficulties, full-scale scores on the SB and WISC are the most highly correlated (range $r=0.73$–0.92; $M=0.81$, standard deviation (SD)$=0.07$), followed by verbal scores (range $r=0.70$–0.82; $M=0.77$, $SD=0.05$) then nonverbal scores (range $r=0.55$–0.79; $M=68$, $SD=0.11$) (Lukens and Hurrell, 1996; Prewett and Matavich, 1992, 1994). Despite these high correlations, significant discrepancies between the scores are reported. Specifically, scores on the SB, including full-scale, verbal, and nonverbal scores, have been reported to be higher in children with ID and those with academic difficulties (Lukens and Hurrell, 1996; Prewett and Matavich, 1994) than scores on the WISC. In one study of children with mild ID, 97% obtained a higher FSIQ score on the SB, 4th ed. than on the WISC-III (average = 8 points higher; range 1–20 points), with 62% of those children showing SB scores that were statistically higher than WISC scores (Lukens and Hurrell, 1996). Verbal and nonverbal scores on the SB were also higher than corresponding scores on the WISC in this sample, with an average difference of about 10 and 5, respectively (Lukens and Hurrell, 1996). Results from these studies are based on group mean differences and do not account for the WISC-SB discrepancy in individual children, a variable that is not reported in most studies.

The SB and Wechsler scales differ on some key variables, as SB does not include tasks assessing processing speed and the SB tasks are reportedly more engaging, leading to children's increased attention and persistence on tasks (Garred and Gilmore, 2009). These and other differences may differentially impact performance on the WISC-IV and the SB-5 in individuals with ASD due to the cognitive and behavioral impairments that are characteristic of the disorder. Therefore, it is possible that results from samples of typically developing youth that show similar performance across the two tests may not generalize to individuals with ASD. Furthermore, if the equivalency of IQ scores on commonly used measures is compromised in children with ASD to the degree it is in children with other clinical diagnoses, such as ID, the validity of drawing inferences from these measures should be questioned.

In the absence of a direct comparison of Wechsler and SB measures in individuals with ASD, it is not clear whether conclusions about relative patterns of performance across subtests or about level of intellectual functioning are reliable across IQ tests. Thus, the aim of this study was to determine the convergent validity of scores derived from the SB-5 and WISC-IV in an ASD sample. Consistent with results of previous studies, it was hypothesized that FSIQ and verbal abilities (Verbal Comprehension Index (VCI)/VIQ) would have the strongest cross-measure correlation coefficients in an ASD group and would be significantly more closely associated than the inter-test nonverbal IQ scores (Perceptual Reasoning Index (NVIQ) or verbal–nonverbal discrepancy scores. Furthermore, we hypothesized that scores would be higher on the SB-5 than on the WISC-IV.

Method

Participants

Participants were 40 adolescents, aged 10 years 0 months to 16 years 11 months ($M=12$ years 11 months, $SD=1$ year 11 months) with a diagnosis of ASD. See Table 1 for additional demographic and clinical information. All participants had previously undergone a multidisciplinary evaluation at a specialized autism center within a large pediatric medical center that included assessments by a developmental pediatrician, clinical psychologist, and speech-language pathologist (SLP). The majority of participants ($n=38$) were administered the Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2000) by research-trained evaluators as part of the standard SLP evaluation. At the time two of the participants were diagnosed, the standard multidisciplinary evaluation did not include the ADOS. However, for all participants, a final clinical diagnosis was conferred by the developmental pediatrician with expertise in ASD, who compiled all assessment results. Based on parent report from all but two families, the average age at their child's first ASD diagnosis was 5 years (range 2–10 years). To confirm that participants had symptoms consistent with an ASD diagnosis at the time of this study, parents completed the Social Responsiveness Scale (SRS) (Constantino and Gruber, 2005). The majority of participants (92.7%) met the recommended cut-off of $T \geqslant 60$ for clinically significant symptoms of ASD; however, three participants obtained T scores below this cut-off—52, 54, and 57. These individuals were retained in the study sample given their prior documentation of an ASD diagnosis.

Eligibility criteria were assessed through a telephone screening. Adolescents with limited verbal communication (i.e. primarily nonverbal or using only single words) or whose parents believed they would be unable to meet the demands of the testing situation (could not point to items when asked, sit in a chair and work at a table, or follow basic commands) were excluded. Additional exclusion criteria included the presence of a genetic or metabolic disorder, a history of neurological illness or injury (e.g. traumatic brain injury, tumor, stroke, epilepsy, Tourette's syndrome), other serious medical conditions that could affect cognition, and severe psychiatric comorbidities such as schizophrenia or bipolar disorder. Individuals who spoke English as a second language, had a history of substance abuse, had significant hearing, vision, or mobility impairments, or

Table 1. Individual demographic and diagnostic data ($n = 40$).

	N	%
Gender		
Male	35	87.5
Female	5	12.5
Race		
White	34	85.0
African American	2	5.0
Native American	1	2.5
Bi-Racial	3	7.5
Diagnosis		
Autism	17	42.5
Asperger's	8	20.0
PDD-NOS	3	7.5
ASD (some clinicians did not specify beyond ASD)	12	30.0

PDD-NOS: pervasive developmental disorder not otherwise specified; ASD: autism spectrum disorder.

who had undergone intelligence testing in the preceding 3 months were also excluded.

Parents or legal guardians of the participants were generally well educated (88% of parents had completed at least some college) and, although there was considerable variability in income (range: <US$20,000 to >US$161,000), as a group they had above-average yearly income (median annual income was between US$66,000 and US$80,000).

Procedures

All aspects of this study were approved by the Institutional Review Boards of the participating hospital and university institutions. The study used a within-subject design to evaluate differences in performance on the WISC-IV and SB-5. As part of a larger battery of neuropsychological measures, these tests were administered in a clinic setting by the first author in a counterbalanced order across participants. Most participants completed the entire assessment battery in less than 3.5 h and all but one participant was able to complete the study in a single appointment (with the remaining individual requiring a second appointment because of fatigue). Participants were rated on key behavioral domains during testing (e.g. attention, cooperation, impulsivity, activity level, motivation) to account for their effect on possible score differences. Most participants had moderate to strong attention, cooperativeness, and motivation, with 5% showing poor cooperation and about 13% showing poor motivation.

Statistical analyses

Complete scores on the WISC-IV were unavailable for four participants. For one of these four participants, the Symbol Search subtest, which is one of two subtests that factors into the Processing Speed Index (PSI), was not administered due to administrator error. Thus, the PSI and the FSIQ could not be calculated (Wechsler, 2003). Three other participants obtained raw scores of 0 on at least two of the three verbal subtests, so a VCI could not technically be calculated (Wechsler, 2003). However, to maximize the number of participants in the between-test comparisons, these three individuals were assigned the lowest possible score of 45 on the VCI. Removal of these three participants in the VCI analyses did not substantially change results.

We assessed convergent validity between and within IQ tests. This was assessed by calculating correlations between pairs of indices from the two intelligence tests that were purported to estimate similar aspects of intelligence, including full-scale, verbal, and nonverbal intelligence, and working memory. We also calculated correlations between different index scores on the same tests and between index scores across tests measuring different constructs. Index scores measuring different constructs on tests (i.e. verbal and nonverbal intelligence, working memory) are correlated in standardization samples and therefore provide further assessment of convergent validity. However, these scores correlate to a lesser degree than scores assessing the same construct and were therefore expected to show weaker correlations in our sample. Index scores from one IQ test that did not have corresponding scores on the other measure, including the WISC-IV PSI and SB-5 Knowledge score were included in correlational analyses, but were not compared. Fisher's r to z transformations were done to facilitate comparisons between correlations.

Verbal–nonverbal discrepancy scores were calculated by subtracting nonverbal scores (PRI and NVIQ) from verbal scores (VCI and VIQ, respectively); thus, a negative number indicated higher nonverbal scores while a positive number indicated higher verbal scores. Although the WISC-IV provides discrepancy score base rates for age and varying IQ levels, the SB-5 provides base rates based on age only. For maximum consistency across measures, only the tests' age group base rates were used to determine the presence or absence of a significant difference between verbal and nonverbal tests. Paired sample t-tests were used to examine mean differences between all corresponding index scores across measures (FSIQ, VCI/VIQ, PRI/NVIQ, Working Memory (WM) Index/WM, and the "Verbal"/"Nonverbal" discrepancy). Differences in IQ classifications across the two tests were assessed by comparing the descriptive IQ classifications (e.g. average, superior) suggested by the test developers using a kappa coefficient. The test order was compared across all index scores to determine whether order of test administration affected performance. Statistical tests were always two-sided with alpha set at 0.05.

Results

Test order

The effect of test order was examined by comparing all IQ and index scores for the test administered first to that

Table 2. Correlations between the Wechsler Intelligence Scale for Children, 4th ed. (WISC-IV) and Stanford–Binet, 5th ed. (SB-5).

	SB.FSIQ	SB.VIQ	SB.NVIQ	SB.WM	SB.FR[a]	SB.KN[a]	SB.QR[a]	SB.VS[a]
WISC.FSIQ	0.88	0.85	0.83	0.79	0.79	0.83	0.80	0.77
WISC.VCI	0.84	0.86	0.73	0.75	0.73	0.85	0.69	0.77
WISC.PRI	0.81	0.74	0.82	0.66	0.78	0.74	0.78	0.77
WISC.WMI	0.77	0.76	0.72	0.78	0.72	0.71	0.71	0.64
WISC.PSI[a]	0.61	0.58	0.60	0.61	0.54	0.58	0.59	0.51

FSIQ: Full-Scale IQ; VIQ: Verbal IQ; VCI: Verbal Comprehension Index; NVIQ: Nonverbal IQ; PRI: Perceptual Reasoning Index, WM/WMI: Working Memory/Working Memory Index; FR: Fluid Reasoning; KN: Knowledge; QR: Quantitative Reasoning; VS: Visual-Spatial; PSI: Processing Speed Index.
[a]Composite scores are listed to show correlations with other scales, but were not evaluated for convergent validity.

Table 3. IQ Data (Wechsler Intelligence Scale for Children, 4th ed. (WISC-IV) and Stanford–Binet, 5th ed. (SB-5)).

	WISC-IV			SB-5			*t*	*d*
	n	*M* (*SD*)	Range	*n*	*M* (*SD*)	Range		
FSIQ	36[a]	78.81 (19.38)	40–121	36[a]	82.06 (19.03)	40–122	−2.05*	−0.17
VCI/VIQ	37[a]	83.49 (20.41)	45–124	37[a]	79.59 (20.09)	43–119	2.29*	0.19
PRI/NVIQ	40	87.55 (20.11)	51–129	40	84.73 (20.36)	43–132	1.47	0.14
WMI/WM	40	76.38 (21.00)	50–123	40	78.78 (18.46)	48–109	−1.14	−0.12
PSI	39[a]	77.79 (15.95)	50–112	–	–	–	–	
Fluid Reasoning	–	–		40	85.68 (20.84)	47–132	–	
Knowledge	–	–		40	79.53 (19.57)	49–111	–	
Quantitative Reasoning	–	–		40	82.40 (19.37)	50–130	–	
Visual Spatial	–	–		40	88.03 (19.57)	50–126	–	

FSIQ: full-scale IQ; VCI: Verbal Comprehension Index; PRI: Perceptual Reasoning Index; WMI: Working Memory Index; PSI: Processing Speed Index; VIQ: Verbal IQ; NVIQ: Nonverbal IQ; WM: Working Memory; *SD*: standard deviation.
[a]Four participants did not have a valid FSIQ; three did not have a valid VCI; and one did not have a PSI score.
*$p < 0.05$.

administered second. The mean differences between scores were 1.69, −0.81, and 2.33 for FSIQ, verbal, and nonverbal scores, respectively. Paired samples *t*-tests revealed no significant differences based on test order for FSIQ, $t(35)=1.03$, $p=0.31$, $d=0.09$; verbal scores, $t(36)=-0.48$, $p=0.67$, $d=-0.04$; or nonverbal scores, $t(39)=-1.10$, $p=0.28$, $d=0.11$. Therefore, the test order was not included in any of the subsequent analyses.

Convergent validity analyses

All correlations presented in Table 2 were significant at $p<0.001$. With regard to convergent validity, it was hypothesized that verbal index correlations would be the strongest across the two tests; however, Fisher's r to z transformations indicated that the verbal score correlations ($r=0.86$) were not significantly different from nonverbal score correlations ($r=0.82$). All indices measuring similar constructs across tests showed strong convergent validity (mean $r=0.84$). The average correlation between index scores measuring different constructs (e.g. WISC-IV VCI and WISC-IV WMI, as well as WISC-IV VCI and SB-5 WM) within and across the two tests ($r=0.71$) was not significantly lower than the other correlations, $p=0.09$.

Full-scale IQ

As anticipated, FSIQ scores on the two measures were highly correlated, $r=0.88$, $p<0.001$. However, a paired samples *t*-test revealed that FSIQ scores on the WISC-IV were significantly lower than FSIQ scores on the SB-5, $t(35)=-2.05$, $p=0.048$, $d=-0.17$ (see Table 3), with an average difference of 3.25 points ($SD=9.51$). Although SB-5 scores were higher on average, individual difference scores ranged from the SB-5 FSIQ being 23 points higher than the WISC-IV to the WISC-IV FSIQ being 22 points higher than the SB-5.

The manuals for the SB-5 and WISC-IV list 95% confidence intervals for the various indices that all fall within 12 IQ points of one another. Thus, a difference of 1 *SD* (15 points) was used in the present analyses to conservatively indicate a clinically significant difference between test scores. Of the 36 participants for whom an FSIQ was calculated on both the WISC-IV and SB-5, the majority (72.2%) scored higher on the SB-5, with four participants scoring at least 1 *SD* higher. Two participants achieved exactly the same FSIQ on the two measures, and only one participant obtained an FSIQ on the WISC-IV that was at least 1 *SD* higher than their score on the SB-5.

Verbal intelligence scores

Verbal intelligence scores on the WISC-IV (VCI) were highly correlated with verbal intelligence scores on the SB-5 (VIQ; $r=0.86$, $p<0.001$). However, VCI scores on the WISC-IV were significantly higher than VIQ scores on the SB-5, $t(36)=2.23$, $p=0.03$; $d=0.19$. As seen in Table 3, the WISC-IV VCI was, on average, 3.89 ($SD=10.62$) IQ points higher than the SB-5 VIQ for the 37 participants with complete data. Nevertheless, most (81.1%) obtained a verbal score on the SB-5 that was within 1 *SD* of their score on the WISC-IV. Differences in scores ranged from the VIQ on the SB-5 being 16 points higher than on the WISC-IV VCI to the VCI being 33 points higher than the VIQ.

Nonverbal intelligence scores

Nonverbal intelligence scores on the WISC-IV (PRI) were strongly correlated with NVIQ scores on the SB-5 ($r=0.82$, $p<0.01$), and scores on the two tests did not differ significantly from one another, $t(39)=1.47$, $p>0.05$, $d=0.14$. Difference scores ranged from 0 to 37 points and occurred in both directions, with the SB-5 NVIQ score being as much as 26 points higher than the PRI for one adolescent, and the PRI being as much as 37 points higher than the NVIQ for another adolescent. Only two participants scored at least 1 *SD* higher on the SB-5 NVIQ, and five scored at least 1 *SD* higher on the WISC-IV PRI.

Working memory scores

Working memory scores on the WISC-IV (Working Memory Index (WMI)) were highly correlated with working memory scores on the SB-5 (WM; $r=0.78$, $p<0.01$) and did not differ significantly from these scores, $t(39)=-1.14$, $p>0.05$, $d=-0.12$. Discrepancies between working memory scores on the two measures ranged from the SB-5 WM score being 27 points higher than the WISC-IV WMI to the WMI being 37 points higher than the WM. Of the sample, 20% ($n=8$) received SB-5 WM scores that were at least 1 *SD* higher than their WISC-IV WMI score, while 12.5% ($n=5$) scored at least 1 *SD* higher on the WISC-IV WMI.

Symptom severity, as measured by SRS scores, did not account for individual cross-measure differences of any of the index scores. Importantly, results did not differ significantly when individuals with SRS scores below the clinical cut-off ($T<60$) were removed.

Classification of intellectual functioning based on index scores across measures

The discrepancies between IQ scores on the two measures were also examined in terms of their implications for categorical classification. FSIQ scores are used to classify individuals into qualitative ranges of cognitive ability designated by the test developers (i.e. superior—IQ $\geqslant$ 120; average—80 $\geqslant$ IQ $\leqslant$ 119; borderline—70 $\geqslant$ IQ $\leqslant$ 79; ID—IQ $<$ 70), and this classification is used both clinically and in research.

All but six participants in the sample (83.3%) were classified into the same descriptive category by both IQ measures, resulting in high agreement, kappa=0.75 (95% confidence interval 0.57–0.93). Of these six individuals, two (5.6% of the total sample) were classified as ID on the WISC-IV but borderline on the SB-5 (with scores of 71 and 70). Two participants (5.6% of total sample) were classified as borderline on the WISC-IV but average on the SB-5. Finally, one participant (2.8% of total sample) was classified as average on the WISC-IV and borderline on the SB-5. Another participant was classified as average on the WISC-IV and superior on the SB-5. Importantly, participants with differences in classification did not show disproportionately more behavioral issues during testing or differential behaviors during administration of the two tests. Figure 1 shows each participant's FSIQ scores across the two measures with lines illustrating cut-offs for ID. For verbal index scores, nine participants were classified differently on the two tests, with most (78%) falling in a higher classification category on the WISC-IV. The NVIQ score classifications differed for 10 participants, with about equal number of participants scoring higher on WISC-IV (60%) as on SB-5 (40%).

Verbal–nonverbal IQ discrepancy scores on the WISC-IV and SB-5

A correlation between verbal–nonverbal discrepancy scores across the two IQ measures revealed that they were significantly related, $r=0.48$, $p<0.01$. A paired samples *t*-test showed that the average verbal–nonverbal discrepancy on the WISC-IV ($M=6.38$, $SD=16.42$), with nonverbal scores being higher than verbal scores, did not differ significantly from the average discrepancy on the SB-5, which also showed nonverbal scores were generally higher than verbal scores, $M=8.00$, $SD=11.21$; $t(36)=0.67$, $p>0.05$.

Significant verbal–nonverbal discrepancies are classified in the ASD literature in a number of ways, including use of test manuals, a 12-point discrepancy (e.g. Ozonoff et al., 2000; Siegel et al., 1996) and a 15-point discrepancy (e.g. Bishop et al., 2011; Chapman et al., 2011). Table 4 illustrates the number of adolescents in the sample who were identified as having a significant verbal–nonverbal IQ discrepancy based on the respective test manuals (Roid, 2003; Wechsler, 2003), as well as the direction (higher verbal vs nonverbal score) of any significant discrepancy.

In total, 18 (48.6%) individuals had significant verbal–nonverbal discrepancies on the WISC-IV and 16 individuals (43.2%) had significant verbal–nonverbal discrepancies on the SB-5, which is comparable to what is seen in other ASD samples (Charman et al., 2011). However, as shown

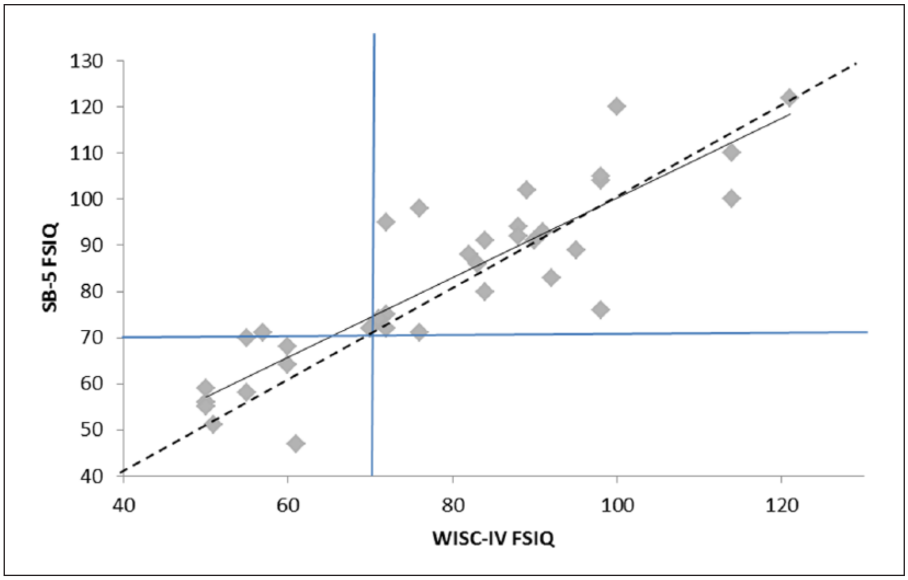

Figure 1. Correlation between full-scale IQ (FSIQ) on the Wechsler Intelligence Scale for Children, 4th ed. (WISC-IV) and Stanford–Binet, 5th ed. (SB-5; $n = 36$).
Dotted line is unity line; cut-off for intellectual disability, IQ < 70, is indicated on both measures.

Table 4. Classification of verbal–nonverbal IQ score discrepancy based on test manuals ($n = 37$).

		Wechsler Intelligence Scale for Children, 4th ed. (WISC-IV)			
		V > NV	V = NV	V < NV	
Stanford–Binet, 5th ed. (SB-5)	V > NV	0 (0%)	2 (5.4%)	0 (0%)	2 (5.4%)
	V = NV	3 (8.1%)	13 (35.2%)	5 (13.5%)	21 (56.8%)
	V < NV	1 (2.7%)	4 (10.8%)	9 (24.3%)	14 (37.8%)
		4 (10.8%)	19 (51.4%)	14 (37.8%)	

V: Verbal Comprehension Index for WISC-IV and Verbal IQ for SB-5; NV: Perceptual Reasoning Index for WISC-IV and Nonverbal IQ for SB-5.

in Table 4, only 59.5% of the sample was classified similarly across the two tests. Only 9 of the 18 who had discrepancies on either test were classified as having the same significant discrepancy classification (e.g. V<NV) on both tests.

Discussion

IQ scores are given substantial weight in the diagnosis and clinical care of individuals with ASD. In ASD research, they are used to select and match participants, to predict outcomes, and as an outcome variable to determine the efficacy of interventions. Equating IQ scores obtained from different tests rests on the assumption that there is strong agreement across measures. However, given characteristics of the ASD population that may affect test taking, it is not necessarily reasonable to assume that these individuals will perform similarly across tests as do typically developing children. To the authors' knowledge, this study is the first to evaluate the relationship between the WISC-IV and SB-5 in any pediatric population and the first to compare any version of the WISC and SB in individuals with ASD.

The results suggest good convergent validity between the current versions of the WISC and SB, with an average correlation between the corresponding indices on the two tests of 0.84. These findings are consistent with reports comparing scores from earlier versions of the WISC and SB in typically developing children (Prewett and Matavich, 1992, 1994; Roid, 2003) and other pediatric populations (Lukens and Hurrell, 1996; Phelps et al., 1988; Prewett and Matavich, 1992, 1994). Thus, the present data suggest that correlations between the two IQ measures are not disproportionately reduced in a population of verbal, healthy individuals with ASD relative to other pediatric populations described in the literature.

The hypothesis that full-scale and VIQ scores would be more highly correlated than nonverbal scores was not supported. As expected, based on the number of tasks that contribute to each index, the highest correlation was for FSIQ and the lowest (albeit still acceptable at $r=0.78$) for working memory. The correlation between the nonverbal indices ($r=0.82$) was substantially higher than the 0.66 correlation between the WISC-III and SB-5 published in the SB-5 manual (Roid, 2003) and higher than the correlations reported in samples of typically developing individuals (Rothlisberg, 1987) and those with attention deficit hyperactivity disorder (ADHD) (Saklofske et al., 1994).

There are no data comparing nonverbal scores on the SB-5 and WISC-IV, so one possible explanation for this difference is that the two tests have become increasingly convergent with each new edition. Another possibility is that the higher nonverbal score correlation in our sample reflects revisions to the nonverbal tasks on the current edition of the WISC, such as the reduced demands on processing speed, greater focus on fluid reasoning, and inclusion of more easy and more difficult items (Wechsler, 2003). Younger children or those with lower IQ scores who comprised previous samples (Lukens and Hurrell, 1996; Prewett and Matavich, 1992, 1994; Rothlisberg, 1987) may have performed better on the SB nonverbal tasks because of the decreased verbal demands relative to the WISC. Finally, cross-measure correlation differences may reflect true developmental differences between the samples. It is possible that verbal adolescents with ASD who are relatively high-functioning perform more similarly across nonverbal tasks than typically developing youth and those with other developmental conditions, supporting the use of NVIQ scores in studies of children with ASD. However, without a comparison control group or data on how the current versions of the WISC and SB relate, we were unable to directly evaluate this possibility.

At the individual level, it is also not possible to conclude whether the difference in scores across the two measures for an adolescent with ASD is similar to what would be found in other adolescent populations. In our sample, individual differences in FSIQ scores were as much as 23 points, and as much as 33 and 37 points for verbal and nonverbal IQ scores, respectively, with at least 1 *SD* difference between corresponding test scores apparent for 12%–33% of the sample. These differences do not appear to be due to symptom severity or behavioral issues during testing. In some cases, weaker FSIQ scores on WISC-IV appeared to result from weak scores on working memory and processing speed, which are constructs that do not both factor into the SB-5 FSIQ. However, more research is needed to determine whether the magnitude and frequency of these score discrepancies are unique to adolescents with ASD or whether they are also characteristic of other pediatric populations. Additionally, it will be important to explore other factors (language, adaptive skills) that may account for individual differences.

This study showed that the majority of adolescents obtained a higher FSIQ score on the SB-5 compared to the WISC-IV. This same pattern has been observed in other studies, including those involving children with ID and those with academic difficulties, although the average difference between the SB and WISC tended to be higher (by about nine IQ points) than in our sample (Lukens and Hurrell, 1996; Prewett and Matavich, 1994). Despite these differences in absolute scores, most adolescents in this study (87.2%) received FSIQ scores on the two measures that were within 1 *SD* of each other.

Classification using standard descriptive categories (i.e. ID, borderline, average, superior) was also generally consistent across measures. Using FSIQ, nearly 85% of the sample received scores on the two measures that placed them in the same cognitive ability category. The six individuals who were classified differently by the two FSIQ scores had differences ranging from 14 to 23 points, with only one participant obtaining a FSIQ on the WISC-IV that was within 1 *SD* of the FSIQ score on the SB-5. For most of these individuals, the shifts between descriptive cognitive ability ranges involved being one descriptive category lower on the WISC-IV compared to the SB-5, which appear to be due to poor working memory and speed of processing. Importantly, results showed high agreement between the WISC-IV and SB-5 in classification of ID in this sample, with only two participants classified as ID on one test but not the other.

There were no significant differences between group-level verbal–nonverbal discrepancy scores on the WISC-IV and SB-5. However, about 40% of participants were classified differently on the two tests using cut-off points set forth in the respective test manuals. A significant verbal–nonverbal discrepancy score pattern in ASD research has been associated with specific neuropathologies, ASD phenotypes, and chromosomal abnormalities (Chapman et al., 2011; Deutsch and Joseph, 2003; Joseph et al., 2002). Given the unacceptably high variability across measures on verbal–nonverbal score differences, researchers and clinicians should exercise caution when interpreting cognitive profiles and purporting their representation as a subtype of ASD.

Limitations and future directions

This study suggests high convergent validity and minimal group differences between the WISC-IV and SB-5 in adolescents with ASD. Importantly, however, our sample was composed of a relatively select group of verbal adolescents with ASD and uncomplicated medical histories who were all capable of completing a test battery that lasted more than 3 h. Also, mean FSIQ scores were in the borderline and low average ranges on the WISC-IV and SB-5, respectively, with only three participants performing at a level too low to obtain a FSIQ on the WISC-IV. These sample characteristics may limit generalizability of the findings to other adolescents with ASD. Although to the authors' knowledge this is the largest study to date that compares WISC and SB in any patient population, the sample size is still small. A larger sample of children and adolescents would provide more statistical power and allow for more complex comparisons between and within various IQ classification groups.

In addition, the absence of a comparative sample prevents conclusions from being drawn about whether the obtained pattern of differences across IQ tests is specific to ASD or is also characteristic of other populations. Although it is reasonable to suspect that the social, communication, and behavioral impairments characteristic of ASD differentially affect test performance, future studies could include samples of typically developing youth and those with other developmental disabilities to determine specificity of this finding. Furthermore, updated normative data comparing the current versions of the WISC and SB are needed. The WISC-IV and SB-5 evaluated in this study are two of the most commonly administered intelligence measures to individuals with ASD, but there are many instruments used in both research and clinical practice that could be compared. Finally, a longitudinal assessment of IQ using multiple measures would assess the relative stability of IQ and allow clinicians to more accurately interpret changes in scores over time when different measures are used.

Funding

This work was supported by the National Institutes of Health (#RC1MH089721; PI: Dr. Catherine Lord).

References

American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Publishing.

Autism Speaks (2010) ATN registry instruments. Available at: http://www.autismspeaks.org/docs/sciencedocs/atn/ATNRegistryInstruments072009.pdf

Begovac I, Begovac B, Majic G, et al. (2009) Longitudinal studies of IQ stability in children with childhood autism—literature survey. *Psychiatria Danubina* 21: 310–319.

Bishop S, Guthrie W, Coffing M, et al. (2011) Convergent validity of the Mullen Scales of Early Learning and the Differential Ability Scales in children with autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities* 116: 331–343.

Black DO, Wallace GL, Sokoloff JL, et al. (2009) Brief report: IQ split predicts social symptoms and communication abilities in high-functioning children with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 39: 1613–1619.

Chapman NH, Estes A, Munson J, et al. (2011) Genome-scan for IQ discrepancy in autism: evidence for loci on chromosomes 10 and 16. *Human Genetics* 129: 59–70.

Charman T, Pickles A, Simonoff E, et al. (2011) IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological Medicine* 41: 619–627.

Constantino J and Gruber C (2005) *Social Responsiveness Scale (SRS).* Los Angeles, CA: Western Psychological Association.

Deutsch DK and Joseph RM (2003) Brief report: cognitive correlates of enlarged head circumference in children with autism. *Journal of Autism and Developmental Disorders* 33: 209–214.

Elliott C (1990) *Differential Ability Scales*. San Antonio, TX: The Psychological Corporation.

Fombonne E (2005) Epidemiology of autistic disorder and other pervasive developmental disorders. *Journal of Clinical Psychiatry* 66: 3–8.

Garred M and Gilmore L (2009) To WPPSI or to Binet, that is the question: a comparison of the WPPSI-III and SB5 with typically developing preschoolers. *Australian Journal of Guidance and Counselling* 19: 104–115.

Goldstein G, Minshew NJ, Allena DN, et al. (2002) High-functioning autism and schizophrenia: a comparison of an early and late onset neurodevelopmental disorder. *Archives of Clinical Neuropsychology* 17: 461–475.

Howlin P (2003) Outcome in high-functioning adults with autism with and without early language delays: implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders* 33: 3–13.

Howlin P, Goode S, Hutton J, et al. (2004) Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry* 45: 212–229.

Huber DH (2008) *Clinical Presentation of Autism Spectrum Disorders in Intellectually Gifted Students*. Ann Arbor, MI: ProQuest (Dissertations and Theses).

Joseph RM, Tager-Flusberg H and Lord C (2002) Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *Journal of Child Psychology and Psychiatry* 43: 807–821.

Klin A, Saulnier CA, Sparrow SS, et al. (2007) Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: the Vineland and the ADOS. *Journal of Autism and Developmental Disorders* 37: 748–759.

Klin A, Saulnier C, Tsatsanis K, et al. (2005) Clinical evaluation in autism spectrum disorders: psychological assessment within a transdisciplinary framework. In: Volkmar F, Rhea P, Klin A, et al. (eds) *Handbook of Autism and Pervasive Developmental Disorders*. Hoboken, NJ: John Wiley & Sons, pp. 772–798.

Lord C, Risi S, Lambrecht L, et al. (2000) The autism diagnostic observation schedule—generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders* 30: 205–223.

Lukens J and Hurrell R (1996) A comparison of the Stanford–Binet IV and the WISC-III with mildly retarded children. *Psychology in the Schools* 33: 24–27.

Mayes SD and Calhoun SL (2003) Analysis of WISC-III, Stanford–Binet:IV, and academic achievement test scores in children with autism. *Journal of Autism and Developmental Disorders* 33: 329–341.

Mottron L (2004) Matching strategies in cognitive research with individuals with high-functioning autism: current practices, instrument biases, and recommendations. *Journal of Autism and Developmental Disorders* 34: 19–27.

Munson J, Dawson G, Sterling L, et al. (2008) Evidence for latent classes of IQ in young children with autism spectrum disorder. *American Journal on Mental Retardation* 113: 439–452.

Ozonoff S, South M and Miller JN (2000) DSM-IV-defined Asperger syndrome: cognitive, behavioral and early history differentiation from high-functioning autism. *Autism* 4: 29–46.

Phelps L, Bell M and Scott M (1988) Correlations between the Stanford-Binet: Fourth edition and the WISC-R with a learning disabled population. *Psychology in the Schools* 25: 380–382.

Prewett P and Matavich M (1992) Mean-score differences between the WISC-R and the Stanford–Binet Intelligence Scale: fourth edition. *Diagnostique* 17: 195–201.

Prewett P and Matavich M (1994) A comparison of referred students' performance on the WISC-III and the Stanford–Binet Intelligence Scale: fourth edition. *Journal of Psychoeducational Assessment* 12: 42–48.

Roid GH (2003) *Stanford–Binet Intelligence Scales*. 5th ed. Itasca, IL: Riverside Publishing.

Rothlisberg BA (1987) Comparing the Stanford–Binet, fourth edition to the WISC-R: a concurrent validity study. *Journal of School Psychology* 25: 193–196.

Saklofske D, Schwean V, Yackulic R, et al. (1994) WISC-III and SB:FE performance of children with attention deficit hyperactivity disorder. *Canadian Journal of School Psychology* 10: 167–171.

Siegel DJ, Minshew NJ and Goldstein G (1996) Wechsler IQ profiles in diagnosis of high-functioning autism. *Journal of Autism and Developmental Disorders* 26: 389–406.

Thorndike RL, Hagen EP and Sattler JM (1986) *Stanford–Binet Intelligence Scales*. 4th ed. Chicago, IL: Riverside Publishing.

Wechsler D (1991) *Wechsler Intelligence Scale for Children*. 3rd ed. San Antonio, TX: The Psychological Corporation.

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

Wechsler D (2003) *Wechsler Intelligence Scale for Children: Technical and Interpretive Manual*. 4th ed. San Antonio, TX: The Psychological Corporation.

White SW, Scahill L, Klin A, et al. (2007) Educational placements and service use patterns of individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 37: 1403–1412.