

Peabody Picture Vocabulary Test: Proxy for Verbal IQ in Genetic Studies of Autism Spectrum Disorder

Kate E. Krasileva¹ · Stephan J. Sanders¹ · Vanessa Hus Bal^{1,2}

Published online: 28 January 2017
© Springer Science+Business Media New York 2017

Abstract This study assessed the utility of a brief assessment (the Peabody Picture Vocabulary Test—4th Edition; PPVT4) as a proxy for verbal IQ (VIQ) in large-scale studies of autism spectrum disorder (ASD). In a sample of 2,420 proband with ASD, PPVT4:IQ correlations were strong. PPVT4 scores were, on average, 5.46 points higher than VIQ; 79% of children had PPVT4 scores within one standard deviation (+/−15) of their VIQ and 90% were similarly classified as having abilities above or below 70 on both measures. Distributions of PPVT4 and VIQ by de novo mutation status were highly similar. These results strongly support the utility of PPVT4 as a proxy for VIQ in large-scale ASD studies, particularly for genetic investigations.

Keywords Autism spectrum disorder · Intelligence tests · Cognition · Language tests

Introduction

In the past decade, there has been substantial progress in the discovery of loci and genes contributing risk for autism spectrum disorder (ASD). Many of these advances have

been made possible by large multi-site studies (Fischbach and Lord 2010) and international consortia (Buxbaum et al. 2012), resulting in several thousand ASD cases, parents, and unaffected sibling controls. While these samples have been sufficient to advance gene discovery by the identification of rare and de novo variants with large effects, it is estimated that identification of common risk alleles, which will generally yield increases in susceptibility less than 1.5-fold, will require tens of thousands of cases (Geschwind and State 2015). The cost of behavioral phenotyping presents a formidable challenge for such large-scale investigations. The focus of this study is to demonstrate the validity of using the Peabody Picture Vocabulary Test—4th Edition (PPVT-4; Dunn and Dunn 2007) as a lower-cost, less time-intensive alternative to traditional cognitive batteries to obtain a proxy for verbal IQ (VIQ).

Aside from the millions of dollars of funding required to ascertain and sequence large samples, considerable resources have also been dedicated to case identification and behavioral phenotyping. The nature of these assessments varies widely, from case confirmation based on cut-offs from caregiver questionnaires to multimodal behavioral batteries administered and interpreted by one or more highly trained clinicians. Studies suggest that behavioral subtyping has a modest effect on genetic homogeneity, thereby doing little to improve power in genome-wide association studies (Chaste et al. 2015). Nonetheless, the collection of a broad array of behavioral measures has allowed for phenotype–genotype analyses that have shown associations between clinical profiles and particular genetic syndromes (e.g., social anxiety in individuals with duplications at 7q11.23; Sanders et al. 2011; Mervis et al. 2015). In particular, a reduction in IQ has been one of the most consistent findings associated with de novo mutations in ASD. Importantly, however, an excess of mutations is also found

Electronic supplementary material The online version of this article (doi:10.1007/s10803-017-3030-7) contains supplementary material, which is available to authorized users.

✉ Vanessa Hus Bal
Vanessa.HusBal@ucsf.edu

¹ University of California San Francisco Department of Psychiatry, 1550 4th St, San Francisco, CA 94158, USA

² STAR Center for ASD & NDD, 401 Parnassus Ave, San Francisco, CA 94143, USA

in individuals with ASD who have average and above average IQs (Sanders et al. 2015). Further analysis of the relationship between IQ, ASD, and genotype is limited by the lack of IQ data for children with ASD in some cohorts and, in most databases, unaffected parents and siblings. The availability of IQ data in parents and siblings would enable us to begin to better understand the relative contributions of de novo and inherited genetic variation to IQ in affected individuals. The necessity of taking into account parental estimates when assessing abnormalities in heritable traits has been demonstrated for head circumference (Chaste et al. 2013). While that study demonstrated that far fewer probands had abnormally large or small head sizes (given parental head size), the ability to compare a proband's measured cognitive ability to what is "expected" given heritability of IQ would help to understand the extent to which risk variants influence aberrations in IQ.

Collection of IQ data presents a challenge to large-scale ASD investigations. Children with ASD span the full IQ range, from superior cognitive ability to profound intellectual disability (ID). Therefore, a single study may implement multiple different cognitive instruments, including use of tests outside the standardized age range to derive estimates for children with ID. For individuals who can complete standard testing, cognitive batteries often require over an hour to be administered by clinical staff with training in psychometrics. Even abbreviated cognitive batteries requiring 30 min may be prohibitive in studies aiming to recruit thousands of participants, particularly those seeking to establish estimates for non-affected family members.

In contrast to cognitive tests, the PPVT-4 is a standardized measure of one-word receptive vocabulary that can be administered in 10–15 min by research staff with minimal training (Dunn and Dunn 2007). This test is available in both paper and digital form, potentially providing further flexibility in administration. Moreover, the PPVT-4 does not necessitate spoken language and has norms for 2–90 year olds, allowing use of the same test across individuals of different ages and varying cognitive abilities.

Since the publication of the first version of the PPVT in 1959, researchers and clinicians have been interested in using this brief assessment as a cognitive screener or proxy for verbal IQ. Early studies of the original PPVT and the Wechsler Adult Intelligence Scale yielded inconsistent results, with some studies suggesting that the PPVT underestimates IQ and others reporting that the PPVT overestimates abilities (Altepeter and Johnson 1989; Carvajal et al. 1989; Craig and Olson 1991; Mangiaracina and Simon 1986; Maxwell and Wise 1984; Price et al. 1990; Prout and Schwartz 1984; Stevenson 1986). Subsequent versions of the PPVT have also been reported to demonstrate moderate to strong correlations with measures of verbal IQ (Dunn and Dunn 1981, 1997). For

example, in a study of typically developing young children of low socio-economic status, Campbell et al. (2001) reported a correlation of 0.58 between PPVT-III and the Kaufman Mental Processing Composite. Whilst PPVT-III scores tended to be lower than the Kaufman (mean difference = 8.3 points), broad IQ classifications (i.e., <70 on both measures) were accurate in 83% of cases. In a small study of adult college students, PPVT-III was moderately correlated (0.46) with VIQ from the Wechsler Adult Intelligence Scale (Bell et al. 2001). Notably, 33% of participants exhibited PPVT-WAIS VIQ differences greater than 10 points; the majority of these cases were PPVT underestimating VIQ for individuals in the superior range (i.e., 120+). Fewer studies have investigated validity of the current PPVT, PPVT-4, though there is some evidence suggesting similar associations with the Leiter-R, a measure of nonverbal cognitive ability, across small samples of typically developing school-aged children and youth with Down Syndrome or idiopathic ID (Phillips et al. 2014).

Although using the PPVT-4 has many advantages over traditional cognitive assessments, its use as a proxy for VIQ (or intelligence more broadly) also carries considerable limitations. Intelligence is a complex construct comprising multiple different cognitive processes. Although there are many competing theories of intelligence (Reschly et al. 2002), most cognitive batteries provide separate estimates for VIQ and nonverbal intelligence (NVIQ). Verbal IQ is generally thought to reflect crystallized intelligence (i.e., based on knowledge and experience). Nonverbal IQ is associated with fluid intelligence, or the capacity to reason about problems, independent of past knowledge. A Full-Scale IQ or General Conceptual Ability score often combines both VIQ and NVIQ, and sometimes other aspects of cognition, such as memory or processing speed. Use of a brief receptive vocabulary measure will not capture potentially discrepant VIQ–NVIQ profiles (e.g., Joseph et al. 2002; Bal et al. 2016). Impairments in cognitive processes such as memory or processing speed are also not assessed using the PPVT-4.

Even within VIQ assessments, there is a combination of abilities assessed, including both expressive and receptive skills. Studies suggest that rates of receptive and expressive vocabulary development may differ for some children with ASD (e.g., Kover et al. 2013). Consequently, as a measure of one-word receptive vocabulary, the PPVT-4 may over- or underestimate VIQ in cases where there are discrepant expressive and receptive abilities. Although specific language profiles are not consistent on a larger group level across studies, there remains the possibility that small subgroups of children show discrepant abilities (Kwok et al. 2015), which would suggest caution is warranted when using the PPVT-4 as a proxy for VIQ.

While there are limitations to any brief assessment, it is important to note that the measurement of IQ in genetic studies (and many ASD studies more broadly) is also affected by methodological limitations. Studies combine IQ estimates from multiple instruments, which may be derived from divergent theories of intelligence assessing different cognitive processes. Nonetheless, there have been consistent associations between broad IQ estimates and genetic mutations. The extraordinarily large sample sizes required of genetic investigations will preclude comprehensive, detailed assessments, particularly when such investigations seek to include family members. The potential for a relatively quicker, although less precise, estimate will enable researchers to answer initial questions (e.g., relative contributions of de novo and inherited variants to IQ as described above). Once a broad understanding of the effect that risk mutations seem to have for IQ is established, more comprehensive evaluations to parse out the specific cognitive or language processes underlying the observed effect can be carried out in smaller genetics-first studies. An example of this is a study by Hippolyte et al. (2016), in which a detailed neuropsychological assessment was used to characterize different cognitive domains affected by CNVs at the 16p11.2 locus. Thus, while there is potential for genotype–phenotype associations to be missed when using proxies for complex phenotypes such as IQ, the costs of these kinds of comprehensive assessments on tens of thousands of probands and family members must be weighed against the proportion expected to have an identified de novo CNV or LoF risk mutation (e.g., 10.5% of simplex cases across all different loci, with any given mutation representing fewer than 1%; Sanders et al. 2015).

The PPVT is commonly used in ASD research, particularly for genetic studies and other large consortia (e.g., Staal et al. 2012; McCray et al. 2014; Sykes et al. 2009). Although an ideal assessment would encompass both verbal and nonverbal abilities, in studies where resources limit the depth of characterization, a proxy for VIQ may be of particular interest given findings suggesting VIQ, but not NVIQ, is a significant predictor of ASD symptom trajectories (Gotham et al. 2012). However, studies using the current edition of the PPVT, PPVT-4, primarily cite studies comparing PPVT-R or PPVT-III with older versions of cognitive assessments to justify its use as a proxy for VIQ. In one recent study, the PPVT-4 was moderately correlated with scores from the Raven Matrices (0.56), a nonverbal IQ assessment, but it was not compared to a measure of VIQ and the focus of the study was minimally verbal children with ASD (Plesa-Skwerer et al. 2015). No study has systematically investigated the convergent validity of the PPVT-4 and other cognitive measures in children with ASD across a wide range of ages and ability levels. Moreover, we are not aware of a study examining how the

PPVT-4:VIQ association may differ when VIQ is estimated using different cognitive assessment instruments in any population. The range of ability levels observed in children with ASD necessitate use of multiple different cognitive tests in large samples, thereby making such examination an important part of validating the PPVT-4 as a proxy for VIQ.

This study specifically adds to the existing literature by (1) systematically investigating the association between PPVT-4 and VIQ assessed using several different traditional cognitive batteries in a large sample of children and adolescents with ASD that span the full range of cognitive abilities; and (2) demonstrating the utility of using PPVT-4 as a proxy for VIQ in genetic studies of ASD.

Methods

Participants

Participants were drawn from the Simons Simplex Collection (SSC), which includes 2,658 children, aged from 4 to 17 years and diagnosed with ASD. In addition to the focus on simplex families (defined as families comprised of one child with ASD who has no known first-, second-, or third-degree relatives with ASD), all children were required to have a nonverbal mental age of at least 18 months, meet ASD cut-offs on the ADOS and ADI-R, and receive a best estimate diagnosis of ASD by a clinician based upon all available information (see Hus et al. 2013). Participants with sensory impairments (e.g., blindness or deafness) that would interfere with standardized testing were excluded. Evaluations were conducted at one of 12 university-based sites; parents completed an ADI-R, Vineland, and questionnaires about the child with ASD. All children completed the ADOS and a cognitive assessment. Study procedures were approved by Institutional Review Boards at each university site. Parents provided informed consent and, when possible, children provided verbal assent. Only those children who completed the PPVT-4 were included in this current analysis (N=2420, 86.4% male, 84.5% White, 88.7% non-Hispanic).

Measures

The Peabody Picture Vocabulary Test—4th Edition (PPVT-4; Dunn and Dunn 2007)

The PPVT-4 offers two alternative forms (A and B), each comprised of 228 items, grouped into 19 sets of 12-items arranged in increasing difficulty. Recommended start points by age or ability level, combined with basal and ceiling rules limit administration time to 10–15 min. Raw scores

are converted to age-normed standard scores with a mean of 100 and standard deviation of 15. The original validation study indicated strong psychometric properties, with split-half reliability and Cronbach's alpha of 0.94 or higher for both forms and test–retest reliability of $r = 0.93$.

Cognitive Level

Verbal (VIQ) and nonverbal IQ (NVIQ) scores were derived from a standard developmental hierarchy of cognitive assessments. Most children received the Differential Ability Scales—2nd edition (DAS-2; Elliott 2007) Early Years and School-Age (87.3% of sample for VIQ and 89.7% for NVIQ). The Mullen Scales of Early Learning (Mullen 1995; VIQ=8.2% and NVIQ=6.2%) or a Wechsler battery (1999, 2003; VIQ=4.1%, NVIQ=4.1%) was used for the remaining sample. In cases where children were unable to obtain a standard IQ estimate on the age-appropriate test, verbal ratio IQs were derived by averaging expressive and receptive age equivalents, then dividing by chronological age and multiplying by 100 (see Bishop et al. 2011). If a child was unable to achieve a basal score on either of the subtests required to derive VIQ, they were administered the next test in the hierarchy and a ratio IQ was derived. The difference between PPVT-4 and VIQ estimate was computed as $PPVT-4 - VIQ$. Discrepancies were defined as differences greater than 15 points (i.e., one standard deviation; $\leq -16 = PPVT < VIQ$, -15 to $15 = PPVT \approx VIQ$, $\geq 16 = PPVT > VIQ$).

Language Level

The Autism Diagnostic Observation Schedule (ADOS; Lord et al. 1999, 2012) is a semi-structured observational assessment used to inform ASD diagnoses. ADOS Module was used as a broad estimate of language level (Module 1: nonverbal or single words, Module 2: phrase speech, and Module 3 or 4: verbally fluent).

Statistical Analyses

Aim 1

Associations between PPVT-4, VIQ, and NVIQ were first examined using Pearson correlations; $r < 0.3$, $0.3-0.5$, and > 0.5 reflect weak, moderate, strong associations, respectively (Cohen 1988). Fisher r -to- z transformation was used to compare correlations across subgroups. There were no differences by sex or race; therefore subsequent analyses were not divided by demographic characteristics. Next, children were divided into four subgroups by age groups reflecting broad developmental levels (preschool:4–5, school-age:6–12, adolescent:13+), VIQ test (DAS-2, Mullen, Wechsler), Cognitive Level (VIQ < 70 , ≥ 70), and Language Level (ADOS Module 1, 2, 3/4). These groupings were chosen to explore whether PPVT-4:VIQ associations differed across subgroups commonly combined in large ASD samples; group descriptives are provided in Table 1. Separate linear regression models predicting VIQ were used to assess whether the

Table 1 Sample descriptives by age, cognitive, and language subgroups

Sample subgroups	N, n	Descriptives [mean (std deviation)]			
		Age	VIQ	NVIQ	PPVT-4
Entire sample	2420	9.08 (3.56)	79.34 (30.23)	85.70 (25.40)	84.80 (29.03)
Age					
4–5 years	541	4.93 (0.58)	80.25 (25.91)	86.99 (22.79)	85.63 (25.03)
6–12 years	1470	8.92 (1.89)	79.76 (29.81)	86.13 (25.22)	85.16 (28.69)
13–17 years	409	15.14 (1.44)	76.64 (36.32)	82.45 (28.85)	82.43 (34.56)
VIQ test					
Mullen	207	7.47 (3.38)	27.64 (15.23)	46.14 (18.16)	39.84 (20.52)
DAS-2	2114	9.08 (3.49)	83.36 (26.60)	88.61 (22.64)	88.16 (25.90)
Wechsler	99	12.46 (3.08)	101.72 (17.56)	106.35 (16.13)	107.11 (20.68)
Cognitive level					
VIQ < 70	781	9.48 (3.61)	43.15 (17.59)	60.35 (20.91)	52.74 (20.55)
VIQ ≥ 70	1639	8.89 (3.52)	96.59 (16.72)	97.78 (17.08)	100.08 (17.87)
Language level					
ADOS Mod 1	399	8.12 (3.51)	36.97 (20.35)	53.55 (21.27)	45.24 (21.59)
ADOS Mod 2	550	7.30 (3.18)	69.28 (22.51)	79.92 (21.44)	75.03 (21.65)
ADOS Mod 3 & 4	1471	10.01 (3.38)	94.60 (21.07)	96.59 (18.85)	99.18 (20.43)

VIQ Verbal IQ, NVIQ Nonverbal IQ, PPVT-4 Peabody Picture Vocabulary Test—4th Edition, DAS-2 Differential Ability Scales—2nd edition, ADOS Autism Diagnostic Observation Schedule, Mod Module

PPVT-4:VIQ association differed across subgroups; predictors were entered in the following order to allow examination of the relative contribution of each set of variables: PPVT-4 score centered at the sample mean, subgroup, and PPVT-4*subgroup interaction. Cohen's f^2 was computed to assess the effects of added predictors, when controlling for all other variables in the model; f^2 of 0.02, 0.15, and 0.35 reflect small, medium, and large effect sizes, respectively (Cohen 1988). PPVT-4 and VIQ distributions were compared; paired t -tests were used to assess individual differences in scores and distributions of PPVT-4–VIQ discrepancies were examined. Finally, PPVT-4 and VIQ categories (i.e., <70 , ≥ 70) were compared to evaluate the impact that large discrepancies had on classification. SPSS 23.0 was used for comparative analyses; given the large sample size and multiple comparisons, the significance level was set at $p < 0.001$.

Aim 2

To demonstrate the utility of PPVT-4 as a proxy for VIQ in genetic studies, PPVT-4 and VIQ distributions were plotted for 447 children with a de novo mutation associated with ASD and 1715 children with no identified copy number variant (CNV) or loss of function (LoF). Data regarding de novo mutations (CNV or LoF) for each child were obtained from Sanders et al., 2015. Details regarding microarray and whole exome sequencing procedures for identifying ASD-associated mutations have been previously reported in detail (see Iossifov et al. 2014 and; Sanders et al. 2015). Consistent with the IQ analyses by Sanders and colleagues (2015), one-sided Wilcoxon Rank Sum Tests (WRST) were used to compare groups; analyses were conducted using R:Version 3.1.1.

Results

Associations Between PPVT-4, VIQ, and NVIQ

Correlations between PPVT-4:VIQ ($r=0.925$), PPVT-4:NVIQ ($r=0.803$), and VIQ:NVIQ ($r=0.824$) were strong. As expected, the PPVT-4:VIQ association was significantly stronger than PPVT-4:NVIQ ($z=25.36$, $p < 0.0001$). Consistent with this, PPVT-4:NVIQ associations within subgroups (i.e., age, cognitive, language) and correspondence between PPVT-4 and NVIQ distributions were reduced (compared to PPVT-4:VIQ comparisons), albeit generally in the same direction. As such, results reported focus on PPVT-4:VIQ; parallel analyses based on NVIQ are available upon request.

Comparisons of PPVT-4:VIQ Associations Across Subgroups

Consistent with the correlations above, PPVT-4 significantly predicted VIQ ($R^2=0.855$). As shown in Table 2, VIQ Test, Cognitive Level, and Language Level also had small, but significant main effects on VIQ ($p < 0.001$), reflecting expected subgroup differences. Only PPVT-4*VIQ Test and PPVT-4:ADOS Module 1 interactions were significant predictors of VIQ ($p < 0.001$), indicating a difference in PPVT-4:VIQ association for different cognitive tests and minimally verbal children. Notably, the change in model R^2 when these interactions were entered into their respective models was less than 0.01, indicating a small effect (i.e., f^2 ; Cohen 1988).

Comparison of PPVT-4 versus VIQ Distributions and Classifications

As shown in Fig. 1a, b, distributions of PPVT-4 and VIQ scores were similar. The PPVT-4 showed a floor effect for 4% of children ($n=100$ at standard score=20), compared to 17% of children ($n=417$) for whom a ratio VIQ was computed due to inability to achieve a standard score on an age-appropriate VIQ test (see Bishop et al. 2011). Ratio PPVT-4 scores could be computed from PPVT-4 age equivalents for 53 cases; when substituted for floor values, PPVT-4 distribution was nearly identical to VIQ (Fig. 1c). Additionally, the PPVT-4 and VIQ both had similarly strong correlations with measures of autism symptomatology, adaptive behavior, and emotional problems (see Supplemental Table 1).

Paired comparisons indicated that PPVT-4 scores were, on average, 5.46 points higher than VIQ, $t(2419)=23.23$, $p < 0.0001$. A closer examination of PPVT-4–VIQ differences revealed that 79% had PPVT-4 scores within one standard deviation (± 15 points) of their VIQ estimate, 18% had PPVT-4 scores exceeding VIQ estimates by 16 or more points, and the remaining 3% had PPVT-4 scores 16+ points below VIQ. Group descriptives are provided in Table 3. VIQ Test, Cognitive Level, and Language Level subgroups showed different proportions of PPVT-4 < VIQ, PPVT-4 \approx VIQ, and PPVT-4 > VIQ. Significance was driven by a higher proportion of PPVT-4 > VIQ in the more cognitively or language impaired groups (Mullen VIQ test, VIQ < 70, ADOS Module 1).

PPVT-4 is plotted against VIQ by classification group in Fig. 2 and distribution of PPVT-4–VIQ differences by classification group is provided in Fig. 3. PPVT-4 and VIQ classifications (i.e., <70 , ≥ 70) were highly consistent; 90% of children in the overall sample had the same PPVT-4 and VIQ classification. Of the 10% misclassified, 9% ($n=208$) were classified as PPVT-4 ≥ 70 and VIQ < 70;

Table 2 Regression models investigating PPVT-4:VIQ associations by subgroup

Sample subgroups	Age				VIQ Test				
	<i>B</i>	<i>SE B</i>	<i>R</i> ²	ΔR^2	<i>B</i>	<i>SE B</i>	<i>R</i> ²	ΔR^2	<i>f</i> ²
Constant	79.421*	0.300*			80.225*	0.239*			
PPVT-4 ^a	0.960*	0.010*	0.855*		0.933*	0.009*	0.855*		
Group			0.855	0.000			0.865*	0.009*	0.074*
4–5 vs. 6–12 years	0.059	0.578		0.000	-25.530*	1.847*			
13+ vs. 6–12 years	-0.419	0.644			6.677*	1.636*			
PPVT-4 × Group Interaction			0.856	0.001			0.870*	0.005*	0.038*
PPVT-4 × 4–5 years	-0.032	0.022		0.007	-0.332*	0.038*			
PPVT-4 × 13+ years	0.035	0.020			-0.269*	0.054*			
Sample subgroups	Cognitive level				Language level				
	<i>B</i>	<i>SE B</i>	<i>R</i> ²	ΔR^2	<i>B</i>	<i>SE B</i>	<i>R</i> ²	ΔR^2	<i>f</i> ²
Constant	85.216*	0.325*			78.386*	0.519*			
PPVT-4	0.744*	0.014*	0.855*		0.932*	0.022*	0.855*		
Group			0.890*	0.035*			0.864*	0.009*	0.066*
VIQ < 70 vs. VIQ ≥ 70	-19.190*	0.739*		0.318*	-10.761*	1.272*			
PPVT-4 × Group Interaction			0.890	0.000			0.865*	0.001*	0.007*
PPVT-4 × VIQ < 70	-0.031	0.022		0.000	3.725*	0.628*			
					-0.157*	0.034*			
					-0.064	0.026			

PPVT-4 Peabody Picture Vocabulary Test—4th Edition, VIQ Verbal IQ, DAS-2 Differential Ability Scales—2nd edition, Mod Module

**p* < 0.0001

^aCentered at sample mean

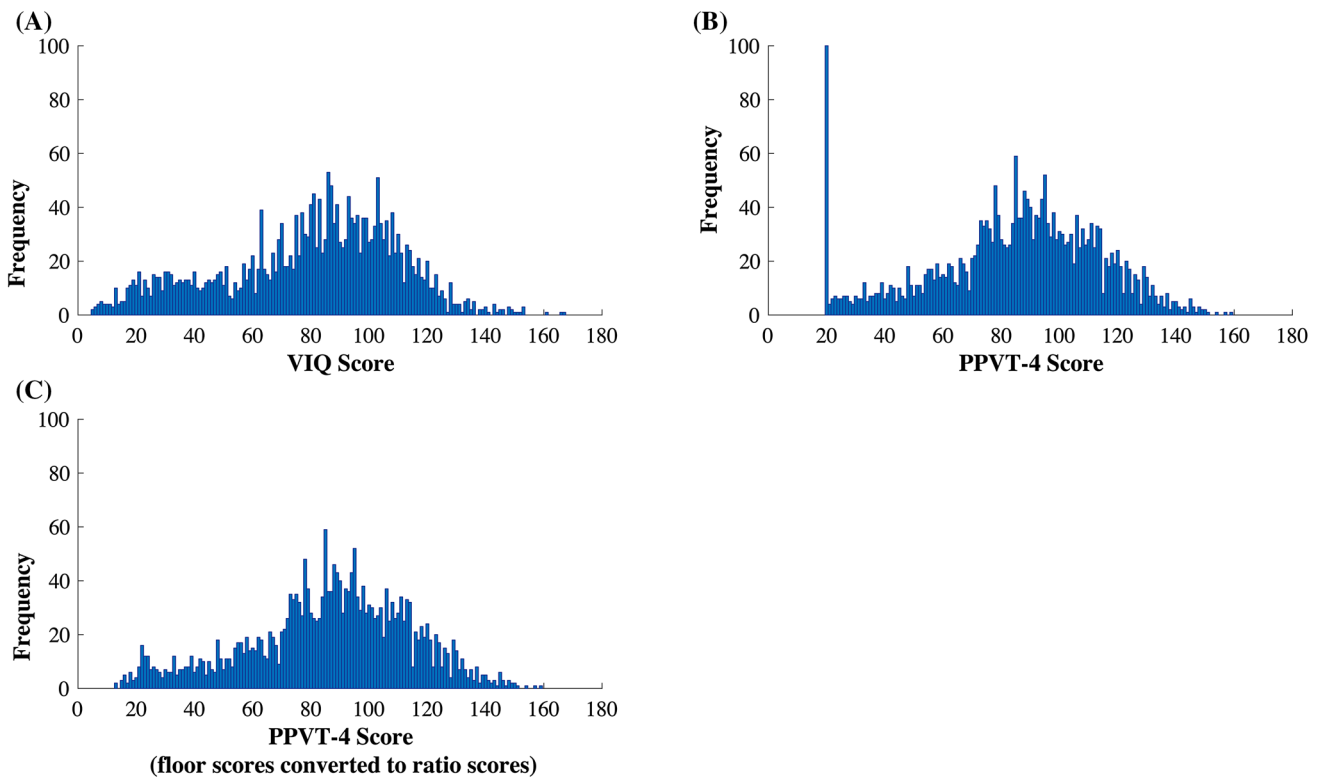


Fig. 1 Comparison of PPVT-4 versus VIQ distributions; *VIQ* Verbal IQ, *PPVT-4* Peabody Picture Vocabulary Test—4th Edition

Table 3 PPVT-4 and VIQ discrepancies across subgroups

Sample subgroups	PPVT-4 < VIQ ^a		PPVT-4 ≈ VIQ ^b		PPVT-4 > VIQ ^c		χ^2
	<i>n</i>	(Subgroup %)	<i>n</i>	(Subgroup %)	<i>n</i>	(Subgroup %)	
Overall	71	(2.9)	1920	(79.4)	429	(17.7)	
Age							
4–5 years	15	(2.8)	433	(80.0)	93	(17.2)	
6–12 years	39	(2.7)	1165	(79.2)	266	(18.1)	2.867
13+ years	17	(4.2)	322	(78.7)	70	(17.1)	
VIQ test							
Mullen	2	(1.0)	138	(66.6)	67	(32.4)	
Wechsler	6	(6.1)	74	(74.7)	19	(19.2)	38.901**
DAS-2	63	(3.0)	1708	(80.8)	343	(16.2)	
VIQ							
<70	11	(1.4)	548	(70.2)	222	(28.4)	95.819**
≥70	60	(3.7)	1372	(83.7)	207	(12.6)	
Language							
ADOS Mod 1	8	(2.0)	301	(75.4)	90	(22.6)	
ADOS Mod 2	7	(1.3)	449	(81.6)	94	(17.1)	17.634**
ADOS Mod 3 & 4	56	(3.8)	1170	(79.5)	245	(16.7)	

PPVT-4 Peabody Picture Vocabulary Test—4th Edition, *VIQ* Verbal IQ, *DAS-2* Differential Ability Scales—2nd Edition, *ADOS* Autism Diagnostic Observation Schedule, *Mod* Module

** $p \leq 0.001$

^aPPVT-4 < VIQ by 16+ points

^bPPVT-4 within 15 points of VIQ

^cPPVT-4 > VIQ by 16+ points

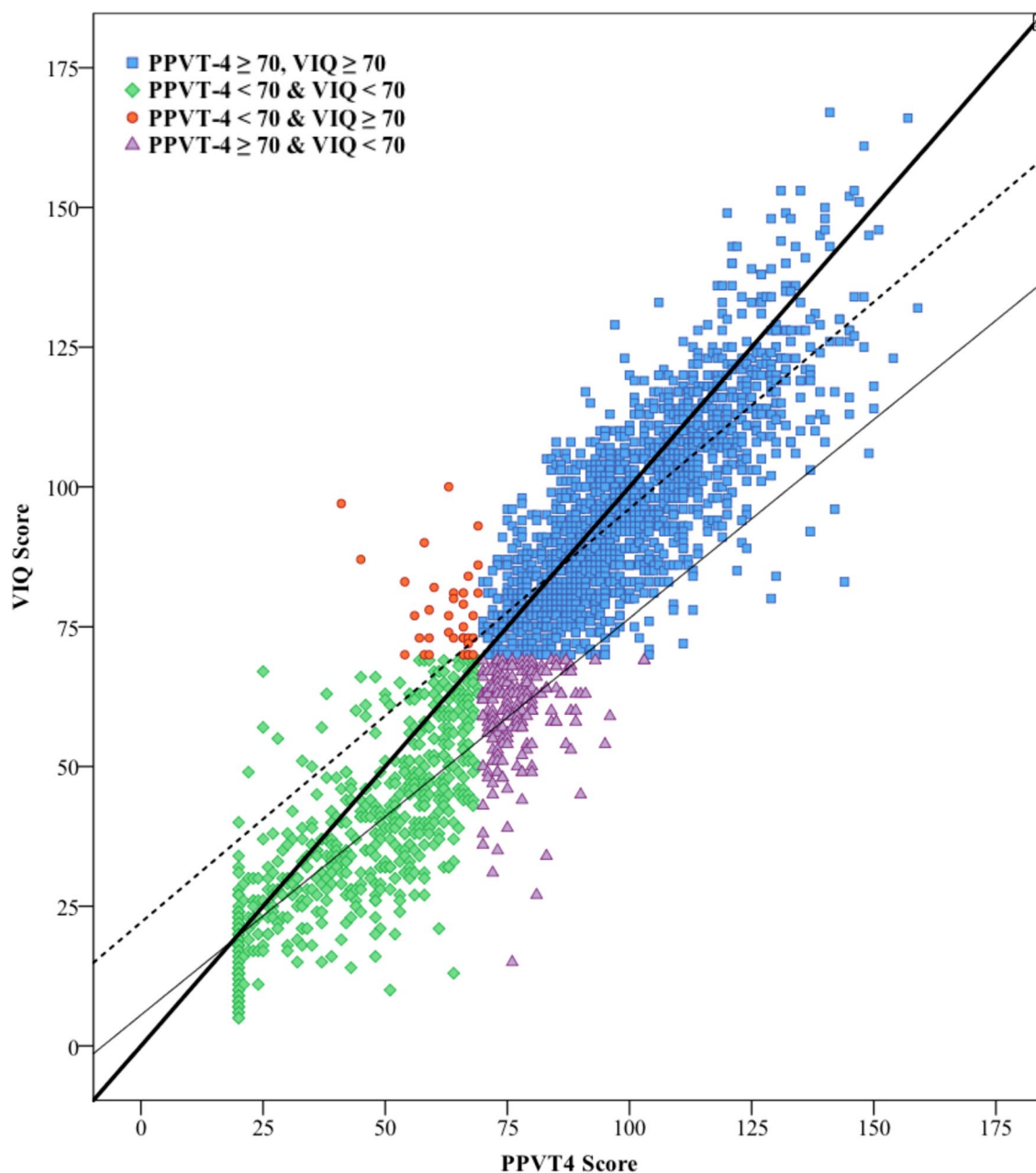


Fig. 2 PPVT-4 versus VIQ by classification group; *PPVT-4* Peabody Picture Vocabulary Test—4th Edition, *VIQ* Verbal IQ. *Thin line* $VIQ < 70$; *thick line* $VIQ = PPVT-4$; *dashed line* $VIQ \geq 70$

49% ($n = 102$) of these were within 15 points (one standard deviation). The remaining 1% ($n = 35$) misclassified were $PPVT-4 < 70$ but $VIQ \geq 70$; the majority of these (57%, $n = 20$) were within 15 points.

Comparison of PPVT-4 and VIQ by de novo mutations

To demonstrate the utility of using the PPVT-4 as a proxy for VIQ, PPVT-4 and VIQ estimates were plotted

by de novo mutation status (Fig. 4). Children with a de novo mutation showed marginally lower VIQ compared to those without an identified mutation (5-point median difference; $p = 0.01$; one-sided WRST); differences in PPVT-4 were also marginal (3-point reduction; $p = 0.04$). Notably, 19 children of the 447 with de novo mutations showed a floor effect on the PPVT-4, compared to 88 children whose verbal abilities were estimated using a ratio VIQ.

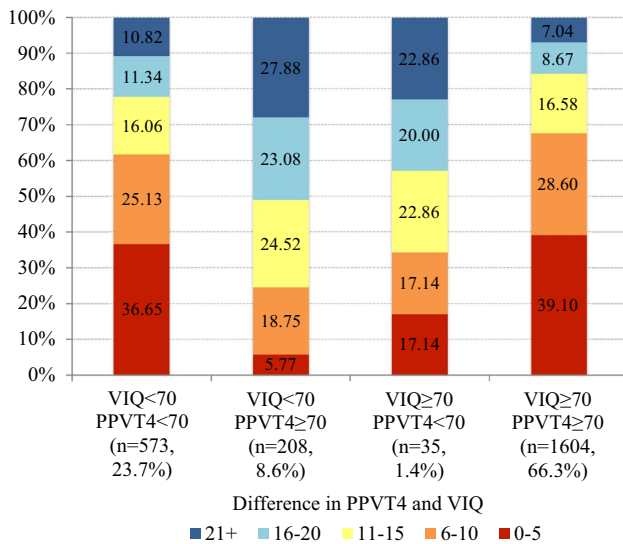


Fig. 3 Distribution of PPVT-4-VIQ differences by classification group; *VIQ* Verbal IQ, *PPVT-4* Peabody Picture Vocabulary Test—4th Edition

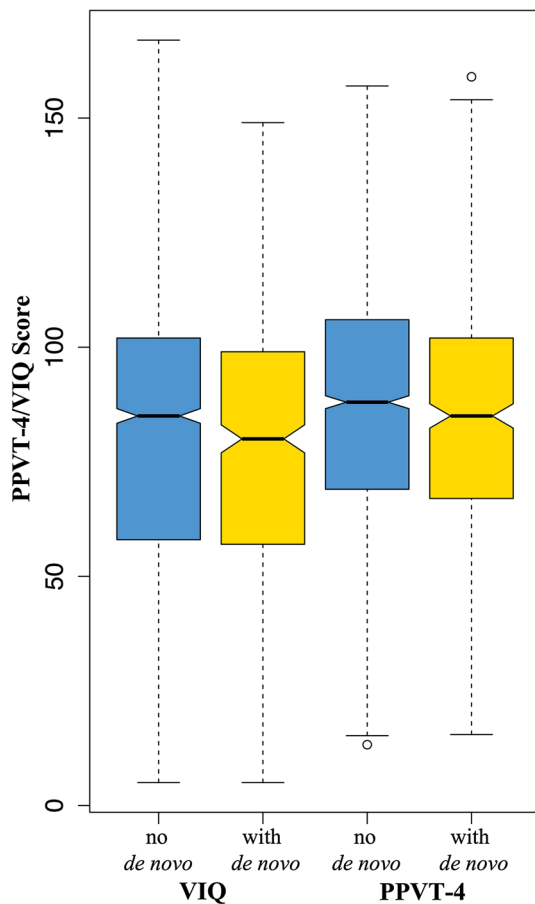


Fig. 4 Comparison of PPVT-4 and VIQ by de novo mutations; *VIQ* Verbal IQ, *PPVT-4* Peabody Picture Vocabulary Test—4th Edition

Discussion

For children with ASD, across varying ages, language, and cognitive ability levels, the PPVT-4 showed strong associations with VIQ estimates derived from different instruments. On average, across groups, the PPVT-4 tended to *overestimate* verbal abilities (by approximately 5.5 points higher than VIQ), with approximately 18% of children exhibiting PPVT-4 scores more than one standard deviation above their estimated VIQ. However, broad cognitive classifications were highly consistent, with only 10% of children misclassified (i.e., $PPVT \geq 70$ when $VIQ < 70$ or vice versa). Importantly, of those 243 children, 26% were within 10 points (the 95% confidence interval range for these cognitive tests; Fig. 3). The distributions of VIQ and PPVT-4 scores by de novo mutation status were nearly identical. Statistical comparisons across groups yielded highly similar results using VIQ or PPVT-4.

While these results support the proposal that the PPVT-4 may be a good proxy for VIQ in large-scale genetic investigations, investigators should be aware of the limitations of the PPVT-4 in lieu of a full cognitive battery. As described in the introduction, IQ is a multi-faceted construct, considered to be comprised of both knowledge gained through experience, as well as more inherent capacity to reason about problems. The DAS-II, used to derive 87% of the VIQ estimates for this study, is modeled after the Cattell–Horn–Carroll theory (CHC; Carroll 1993). Verbal subtests map onto crystallized intelligence (Gc), and nonverbal subtests are further divided into fluid reasoning (Gf) and visual perception (Gv; Elliott 2007). The PPVT-4 is also used as an estimate of crystallized intelligence (Akshoomoff et al. 2013); accordingly, we suggest the PPVT-4 as a potential proxy for VIQ. Analyses indicated that PPVT-4:NVIQ associations were weaker than PPVT-4:VIQ (available upon request), which may reflect separation of crystallized intelligence versus fluid or spatial reasoning (consistent with CHC theory). This suggests that analyses using the PPVT-4 may miss genotype–phenotype associations if mutations are more specifically affecting fluid or spatial reasoning aspects of intelligence (e.g., Edlmann et al. 2007). Given research demonstrating that some individuals with ASD exhibit uneven verbal–nonverbal profiles (e.g., Joseph et al. 2002), it is preferable to obtain separate verbal and nonverbal cognitive estimates whenever feasible. Many studies implementing the PPVT-4 as a proxy for VIQ have used a brief NVIQ test, such as Raven’s Progressive Matrices (Raven 2000). The Matrix Adaptive Test (Hansen 2016) may provide an efficient online alternative to the Raven that awaits further validation.

A second consideration is how different language profiles may affect PPVT-4:VIQ correspondence. For minimally verbal children and/or those with VIQ below 70,

the DAS-Early Years and Mullen were the most common assessments used to derive VIQ. Both tests assess expressive (emphasizing verbal labeling) and receptive language (responding to verbal instructions). The PPVT-4 has lower test demands, in that it does not require spoken language and only assesses single word understanding. Thus, our finding that PPVT-4 tended to be higher than VIQ for children who were minimally verbal, had a $VIQ < 70$ and/or were assessed using the Mullen may be expected, given that VIQ would reflect an average of expressive and receptive abilities. A closer look at the language profiles of individuals with discrepant PPVT-4:VIQ scores may be of interest for more in-depth characterization studies. Although the subgroup of children with discrepant receptive and expressive language abilities may be small (Kwok et al. 2015), considering discrepancies within individual phenotypic profiles may highlight phenotype–genotype associations that would be missed by focusing on broader dimensions, such as VIQ.

Although there is a logical explanation for why PPVT-4 may be overestimating verbal abilities, it is also possible that cognitive tests underestimate VIQ in more impaired children. Considering challenges of assessing children with limited language, it is plausible that the PPVT-4 actually provides a better assessment of their abilities. That the PPVT-4 does not require verbal responses offers a distinct advantage compared to standard cognitive tests, as well as other brief assessments such as, the Expressive Vocabulary Test, Second Edition (EVT-2; Williams 1997). Indeed, the PPVT-4 is one of the few measures recommended as being well-suited for assessing minimally verbal children (Kasari et al. 2013).

Reduced precision is an expected trade-off of using a 10–15 min assessment administered by research staff with minimal training in place of a 30–60+ minute assessment administered by an experienced clinician. However, the cost:benefit ratio of using the PPVT-4 in different types of research contexts needs to be considered for each individual study. Although time and cost are always pertinent factors, whether the PPVT-4 provides an adequate cognitive estimate will vary depending on the purpose of the study and would not be advised in clinical studies which are particularly interested in understanding complex behavioral profiles of subgroups of children. For example, in samples of individuals with intellectual disability, case-control matching on PPVT-4 versus a nonverbal measure is likely to produce different results (Phillips et al. 2014).

Such a trade-off may be less problematic in large-scale genetic studies, where there is a need to have a cognitive estimate for both children with ASD and their parents and/or siblings. In addition to requiring less time and advanced training to administer, an advantage of the PPVT-4 over traditional cognitive assessments is the availability of

standard scores on the same test across a wide range of ages and ability levels. Compared to traditional cognitive assessments, for which standard scores only go as low as 30–50, the PPVT-4 extends down to 20, thereby yielding fewer floor effects. Indeed, only 4% (100/2420) of children were at or below the floor on the PPVT-4, compared to 17% of children for whom ratio VIQs had to be computed due to scores falling below the standard floor or skill level necessitating use of a test outside its standardized age range. While it is common practice to derive ratio IQs for children with ASD who cannot achieve standard scores on age-appropriate cognitive batteries, differences in subtests across tests used for different ages and age effects on calculations make it difficult to compare ratio scores across individuals of varying ages (Bishop et al. 2015). In addition to increasing comparability of children with ASD across varying ability levels, PPVT-4 standard estimates from the same test administered to unaffected siblings and parents will be more easily compared. Such comparability is especially pertinent in genetic studies when the goal is to provide familial context in order to facilitate interpretation of the relative contribution of *de novo* mutations to a proband's IQ.

Regardless of how comprehensive the assessment, behavioral phenotypes, including IQ, are not static within individuals across the lifespan (e.g., Bishop et al. 2015). Thus, it is difficult to compare measurements at a single time point across samples for wide age and developmental ranges encompassed in large genetic studies. Exploration of developmental trajectories within subgroups of individuals with specific ASD risk mutations is more likely to inform associations between biological risk factors and phenotypic outcomes (Lord et al. 2015). Moreover, the complex biological consequences downstream of any given genetic mutation that go on to influence brain development and function are matched by an equally complex interaction of processes, many of which are difficult to capture with behavioral measurement. As such, there is also a need to consider biological markers for more complex behaviors that may be more objectively measured, such as physiological reactivity or neural functioning.

Taken together, brief assessments as proxy estimates of more complex behavioral constructs may be a useful step forward to address the specific question posed by large-scale genetic investigations: what are the relative contributions of *de novo* versus inherited variation to IQ? There will be a need for in-depth, genetic-first longitudinal follow-up studies that take a more comprehensive, multimodal assessment approach to understand specific genotype–phenotype associations. An important limitation to this study relates to the ascertainment of the sample. SSC inclusion criteria required that all children have a nonverbal mental age of at least 18 months and that a direct assessment of VIQ and

NVIQ be obtained for all participants. Thus, it is possible that the proportion of children able to achieve a standard PPVT-4 score (96%) is somewhat higher than might be observed in other samples. Interestingly, researchers have begun to explore ways to use receptive vocabulary measures similar to the PPVT-4 to assess children with ASD; for example, some researchers reduce motor demands or other potential confounds to tests requiring pointing responses by using eye tracking to assess performance (Brady et al. 2014; Plesa-Skwerer et al. 2015). Such methods may actually increase PPVT-4 completion rates and await more systematic validation in larger samples.

In summary, these data suggest that the PPVT-4 is a good proxy for VIQ in children with ASD, providing standard scores across a wide range of age and ability levels. The PPVT-4 has three main advantages over standard cognitive batteries: (1) it is faster (10–15 vs. 30–60+ minutes); (2) it can be administered by research staff with minimal training; and (3) it is comparable across different ages and levels of ability. In addition, it is now available in both digital and paper formats, potentially creating an avenue for siblings or parents to complete it remotely, which would further facilitate large-scale data collection. However, future studies are needed to investigate measurement equivalence across formats and modes of administration. Although there are clear limitations to using only the PPVT-4 as a proxy for cognitive abilities (e.g., it may somewhat overestimate VIQ and is less strongly correlated with NVIQ), this may be an acceptable trade-off in order to expand collection of cognitive estimates to parents or siblings in large-scale studies, such as genetic consortiums aiming to ascertain tens of thousands of families.

Acknowledgments We gratefully acknowledge the SSC families and principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, D. Grice, A. Klin, D. Ledbetter, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC dataset described in this study [v15] by applying at <https://base.sfari.org>.

Authors' Contributions VHB conceived of the study, participated in the design and coordination of the study, lead analysis and interpretation of data and drafted the manuscript; KK participated in the design and coordination of the study, performed statistical analysis and drafted the manuscript; SS participated in the design of the study, interpretation of the data and drafted the manuscript; All authors read and approved the final manuscript.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964

Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

- Akshoomoff, N., Beaumont, J. L., Bauer, P. J., Dikmen, S., Gershon, R., Mungas, D., Slotkin, J., Tulskey, D., Weintraub, S., Zelazzo, P., & Heaton, R. K. (2013). NIH toolbox cognitive function battery (CFB): Composite scores of crystallized, fluid, and overall cognition. *Monographs of the Society for Research in Child Development, 78*, 119–132.
- Altepeter, T. S., & Johnson, K. A. (1989). Use of the PPVT-R for intellectual screening with adults: A caution. *Journal of Psychoeducational Assessment, 7*, 39–45.
- Bal, V. H., Katz, T., Bishop, S. L., & Krasileva, K. (2016). Understanding definitions of minimally verbal across instruments: Evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 57*, 1424–1433.
- Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the peabody picture vocabulary test—Third edition and Wechsler adult intelligence scale—Third edition with university students. *Journal of Clinical Psychology, 57*, 417–422.
- Bishop, S. L., Farmer, C., & Thurm, A. (2015). Measurement of non-verbal IQ in autism spectrum disorder: Scores in young adulthood compared to early childhood. *Journal of Autism and Developmental Disorders, 45*, 966–974.
- Bishop, S. L., Guthrie, W., Coffing, M., & Lord, C. (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.
- Brady, N. C., Anderson, C. J., Hahn, L. J., Obermeier, S. M., & Kapa, L. L. (2014). Eye tracking as a measure of receptive vocabulary in children with autism spectrum disorders. *Augmentative and Alternative Communication, 30*, 147–159.
- Buxbaum, J. D., Daly, M. J., Devlin, B., Lehner, T., Roeder, K., State, M. W., & Autism Sequencing Consortium (2012). The autism sequencing consortium: Large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron, 76*, 1052–1056.
- Campbell, J. M., Bell, S. K., & Keith, L. K. (2001). Concurrent validity of the peabody picture vocabulary test—Third edition as an intelligence and achievement screener for low SES African American children. *Assessment, 8*, 85–94.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carvajal, H., Shaffer, C., & Weaver, K. A. (1989). Correlations of scores of maximum security inmates of Wechsler adult intelligence scale—Revised and peabody picture vocabulary test—Revised. *Psychological Reports, 65*, 268–270.
- Chaste, P., Klei, L., Sanders, S. J., Hus, V., Murtha, M. T., Lowe, J. K., Willsey, A. J., Moreno-De-Luca, D., Yu, T. W., Fombonne, E., Geschwind, D., Grice, D. E., Ledbetter, D. H., Mane, S. M., Martin, D. M., Morrow, E. M., Walsh, C. A., Sutcliffe, J. S., Martin, C. L., Beaudet, A. L., Lord, C., State, M. W., Cook, E. H. Jr., & Devlin, B. (2015). A genome-wide association study of autism using the simons simplex collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? *Biological Psychiatry, 77*, 775–784.

- Chaste, P., Klei, L., Sanders, S. J., Murtha, M. T., Hus, V., Lowe, J. K., Willsey, A. J., Moreno-De-Luca, D., Yu, T. W., Fombonne, E., Geschwind, D., Grice, D. E., Ledbetter, D. H., Lord, C., Mane, S. M., Martin, C. L., Martin, D. M., Morrow, E. M., Walsh, C. A., Sutcliffe, J. S., State, M. W., Devlin, B., Cook, E. H. Jr., & Kim, S. (2013). Adjusting head circumference for covariates in autism: Clinical correlates of a highly heritable continuous trait. *Biological Psychiatry*, *74*(8), 576–584.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Laurence Erlbaum Associates.
- Craig, R. J., & Olson, R. E. (1991). Relationship between Wechsler scales and peabody picture vocabulary test—Revised scores among disability applicants. *Journal of Clinical Psychology*, *47*, 420–429.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Bloomington, MN: Pearson Assessments.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody picture vocabulary test—Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test, third edition*. Circle Pines, MN: AGS.
- Edelmann, L., Prosnitz, A., Pardo, S., Bhatt, J., Cohen, N., Lauriat, T., Ouchanov, L., González, P. J., Manghi, E. R., Bondy, P., Esquivel, M., Monge, S., Delgado, M. F., Splendore, A., Francke, U., Burton, B. K., & McInnes, L. A. (2007). An atypical deletion of the Williams-Beuren syndrome interval implicates genes associated with defective visuospatial processing and autism. *Journal of Medical Genetics*, *44*(2):136–143.
- Elliott, C. D. (2007). *Differential ability scales—Second edition (DAS-II)*. San Antonio, TX: The Psychological Corporation.
- Fischbach, G. D., & Lord, C. (2010). The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron*, *68*, 192–195.
- Geschwind, D. H., & State, M. W. (2015). Gene hunting in autism spectrum disorder: On the path to precision medicine. *Lancet Neurology*, *14*, 1109–1120.
- Gotham, K., Pickles, A., & Lord, C. (2012). Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*, *130*, e1278–e1284.
- Hansen, J. A. (2016). Development and psychometric evaluation of the Hansen Research Services Matrix Adaptive Test: A measure of nonverbal IQ. *Journal of Autism and Developmental Disorders*. doi: 10.1007/s10803-016-2932-0.
- Hippolyte, L., Maillard, A. M., Rodriguez-Herreros, B., Pain, A., Martin-Brevet, S., Ferrari, C., Conus, P., Macé, A., Hadjikhani, N., Metspalu, A., Reigo, A., Kolk, A., Männik, K., Barker, M., Isidor, B., Le Caignec, C., Mignot, C., Schneider, L., Mottron, L., Keren, B., David, A., Doco-Fenzy, M., Gérard, M., Bernier, R., Goin-Kochel, R. P., Hanson, E., Snyder, L. G., 16p11.2 European Consortium, Simons Variation in Individuals Project Consortium, Ramus, F., Beckmann, J. S., Draganski, B., Reymond, A., & Jacquemont, S. (2016). The number of genomic copies at the 16p11.2 locus modulates language, verbal memory, and inhibition. *Biological Psychiatry*, *80*, 129–139.
- Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry*, *54*, 216–224.
- Iossifov, I., O’Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., Stessman, H. A., Witherspoon, K. T., Vives, L., Patterson, K. E., Smith, J. D., Paeper, B., Nickerson, D. A., Dea, J., Dong, S., Gonzalez, L. E., Mandell, J. D., Mane, S. M., Murtha, M. T., Sullivan, C. A., Walker, M. F., Waqar, Z., Wei, L., Willsey, A. J., Yamrom, B., Lee, Y., Grabowska, E., Dalkic, E., Wang, Z., Marks, S., Andrews, P., Leotta, A., Kendall, J., Hakker, I., Rosenbaum, J., Ma, B., Rodgers, L., Troge, J., Narzisi, G., Yoon, S., Schatz, M. C., Ye, K., McCombie, W. R., Shendure, J., Eichler, E. E., State, M. W., & Wigler, M. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, *515*, 216–221.
- Joseph, R. M., Tager-Flusberg, H., & Lord, C. (2002). Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *43*, 807–821.
- Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*, *6*, 479–493.
- Kover, S. T., McDuffie, A. S., Hagerman, R. J., & Abbeduto, L. (2013). Receptive vocabulary in boys with autism spectrum disorder: Cross-sectional developmental trajectories. *Journal of Autism and Developmental Disorders*, *43*, 2696–2709.
- Kwok, E. Y. L., Brown, H. M., Smyth, R. E., & Oram Cardy, J. (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders*, *9*, 202–222.
- Lord, C., Bishop, S., & Anderson, D. (2015). Developmental trajectories as autism phenotypes. *American Journal of Medical Genetics Part C*, *169*, 198–208.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule: ADOS-2*. Los Angeles, CA: Western Psychological Services.
- Lord, C., Rutter, M., DiLavore, P. S., & Risi, S. (1999). *Autism diagnostic observation schedule (ADOS)*. Los Angeles, CA: Western Psychological Services.
- Mangiaracina, J., & Simon, M. J. (1986). Comparison of the PPVT-R and WAIS-R in state hospital psychiatric patients. *Journal of Clinical Psychology*, *42*, 817–820.
- Maxwell, J. K., & Wise, F. (1984). PPVT IQ validity in adults: A measure of vocabulary, not of intelligence. *Journal of Clinical Psychology*, *40*, 1048–1053.
- McCray, A. T., Trevvett, P., & Frost, H. R. (2014). Modeling the autism spectrum disorder phenotype. *Neuroinformatics*, *12*, 291–305.
- Mervis, C. B., Klein-Tasman, B. P., Huffman, M. J., Velleman, S. L., Pitts, C. H., Henderson, D. R., Woodruff-Borden, J., Morris, C. A., & Osborne, L. R. (2015). Children with 7q11.23 duplication syndrome: Psychological characteristics. *American Journal of Medical Genetics Part A*, *167*, 1436–1450.
- Mullen, E. (1995). *The Mullen scales of early learning*. Circle Pines, MN: American Guidance Service, Inc.
- Phillips, B. A., Loveall, S. J., Channell, M. M., & Connors, F. A. (2014). Matching variables for research involving youth with Down syndrome: Leiter-R versus PPVT-4. *Research in Developmental Disabilities*, *35*, 429–438.
- Plesa-Skwerer, D. P., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2015). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism*, *20*, 591–604.
- Price, D. R., Herbert, D. A., Walsh, M. L., & Law, J. G. Jr. (1990). Study of WAIS-R, quick test and PPVT IQs for neuropsychiatric patients. *Perceptual and Motor Skills*, *70*, 1320–1322.
- Prout, H. T., & Schwartz, J. F. (1984). Validity of the peabody picture vocabulary test—Revised with mentally retarded adults. *Journal of Clinical Psychology*, *40*, 584–587.
- Raven, J. (2000). The Raven’s progressive matrices: change and stability over culture and time. *Cognitive Psychology*, *41*, 1–48. doi:10.1006/cogp.1999.0735.
- Reschly, D. J., Myers, T. G., & Hartel, C. R. (2002). *Mental retardation: Determining eligibility for social security benefits*. Washington, D.C.: National Academies Press.
- Sanders, S. J., Ercan-Sencicek, A. G., Hus, V., Luo, R., Murtha, M. T., Moreno-De-Luca, D., Chu, S. H., Moreau, M. P., Gupta, A. R., Thomson, S. A., Mason, C. E., Bilguvar, K., Celestino-Soper, P. B. S., Choi, M., Crawford, E. L., Davis, L., Davis Wright, N.

- R., Dhodapkar, R. M., DiCola, M., DiLullo, N. M., Fernandez, T. V., Fielding-Singh, V., Fishman, D. O., Frahm, S., Garagaloian, R., Goh, G. S., Kammela, S., Klei, L., Lowe, J. K., Lund, S. C., McGrew, A. D., Meyer, K. A., Moffat, W. J., Murdoch, J. D., O'Roak, B. J., Ober, G. T., Pottenger, R. S., Raubeson, M. J., Song, Y., Wang, Q., Yaspan, B. L., Yu, T. W., Yurkiewicz, I. R., Beaudet, A. L., Cantor, R. M., Curland, M., Grice, D. E., Günel, M., Lifton, R. P., Mane, S. M., Martin, D. M., Shaw, C. A., Sheldon, M., Tischfield, J. A., Walsh, C. A., Morrow, E. M., Ledbetter, D. H., Fombonne, E., Lord, C., Martin, C. L., Brooks, A. I., Sutcliffe, J. S., Cook, E. H. Jr., Geschwind, D., Roeder, K., Devlin, B., & State, M. W. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*, *70*, 863–885.
- Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., Murtha, M. T., Bal, V. H., Bishop, S. L., Dong, S., Goldberg, A. P., Jinlu, C., Keaney, J. F., Lambertus, K., Mandell, J. D., Moreno-De-Luca, D., Poultney, C. S., Robinson, E. B., Smith, L., Solli-Nowlan, T., Su, M. Y., Teran, N. A., Walker, M. F., Werling, D. M., Beaudet, A. L., Cantor, R. M., Fombonne, E., Geschwind, D. H., Grice, D. E., Lord, C., Lowe, J. K., Mane, S. M., Martin, D. M., Morrow, E. M., Talkowski, M. E., Sutcliffe, J. S., Walsh, C. A., Yu, T. W., Autism Sequencing Consortium, Ledbetter, D. H., Martin, C. L., Cook, E. H., Buxbaum, J. D., Daly, M. J., Devlin, B., Roeder, K., & State, M. W. (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 Risk Loci. *Neuron*, *87*, 1215–1233.
- Staal, W. G., de Krom, M., & de Jonge, M. V. (2012). Brief report: the dopamine-3-receptor gene (DRD3) is associated with specific repetitive behavior in autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders*, *42*, 885–888.
- Stevenson, J. D. (1986). Alternate form reliability and concurrent validity of the PPVT-R for referred rehabilitation agency adults. *Journal of Clinical Psychology*, *42*, 650–653.
- Sykes, N. H., Toma, C., Wilson, N., Volpi, E. V., Sousa, I., Pagnamenta, A. T., Tancredi, R., Battaglia, A., Maestrini, E., Bailey, A. J., Monaco, A. P., & International Molecular Genetic Study of Autism Consortium (IMGSAC) (2009). Copy number variation and association analysis of SHANK3 as a candidate gene for autism in the IMGSAC collection. *European Journal of Human Genetics*, *17*, 1347–1353.
- Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2003). *Wechsler intelligence scale for children—WISC-IV*. San Antonio, TX: The Psychological Corporation.
- Williams, K. T. (1997). Expressive vocabulary test—Second edition (EVT™ 2). *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*, 864–872.