Research in Autism Spectrum Disorders 8 (2014) 1434–1442

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

Journal homepage: http://ees.elsevier.com/RASD/default.asp

# A comparison of measures for assessing the level and nature of intelligence in verbal children and adults with autism spectrum disorder

Kimberly E. Bodner [a,1,*], Diane L. Williams [a,b,1,2], Christopher R. Engelhardt [c,3], Nancy J. Minshew [a,d,1]

[a] NIH Autism Center of Excellence, University of Pittsburgh School of Medicine, 3811 O'Hara Street, Suite 300, Webster Hall, Pittsburgh, PA 15213-2593, United States

[b] Speech-Language Pathology, Rangos School of Health Sciences, Fisher Hall 409, Duquesne University, 600 Forbes Avenue, Pittsburgh, PA 15282, United States

[c] Department of Health Psychology, Thompson Center for Autism & Neurodevelopmental Disorders, University of Missouri, 205 Portland Street, Columbia, MO 65211, United States

[d] University of Pittsburgh School of Medicine, Departments of Psychiatry & Neurology, 3811 O'Hara Street, Suite 300, Webster Hall, Pittsburgh, PA 15213-2593, United States

ARTICLE INFO

*Article history:*
Received 8 May 2014
Received in revised form 19 July 2014
Accepted 20 July 2014

*Keywords:*
Autism spectrum disorder
High-functioning
Intelligence
Wechsler
Raven's Progressive Matrices
Cognition

ABSTRACT

Previous work has suggested that the Raven's Progressive Matrices (RPM) are better suited for capturing the nature of intelligence for individuals with autism spectrum disorder (ASD) than the Wechsler scales. The RPM measures 'fluid intelligence', an area for which it has been argued that persons with ASD have a relative strength. Given that measures of intelligence are used for establishing clinical diagnoses, for making educational decisions, and for group-matching in research studies, continued examination of this contention is warranted. In the current study, verbal children with ASD performed moderately better on the RPM than on the Wechsler scales; children without ASD received higher percentile scores on the Wechsler than on the RPM. Adults with and without ASD received higher percentile scores on the Wechsler than the RPM. Results suggest that the RPM and Wechsler scales measure different aspects of cognitive abilities in verbal individuals with ASD. For the verbal children and adults with ASD in the current study, the RPM and Wechsler scales have unique contributions that must be considered in context when establishing a baseline of cognitive function. The results of this investigation highlight the importance of thoughtfully selecting appropriate measures of intelligence consistent with clinical, educational, and research purposes, especially for verbal children and adults with ASD.



* Corresponding author. Present address: Department of Psychological Sciences, University of Missouri, 3 McAlester Hall, Columbia, MO 65211, United States. Tel.: +1 573 884 8109; fax: +1 573 882 7710.

*E-mail addresses:* kimberly.bodner@mail.mizzou.edu, kimberly.bodner@gmail.com (K.E. Bodner), williamsd2139@duq.edu (D.L. Williams), engelhardtc@health.missouri.edu (C.R. Engelhardt), minshewnj@upmc.edu (N.J. Minshew).

[1] Tel.: +1 412 246 5460; fax: +1 412 246 5470.
[2] Tel.: +1 412 396 4217; fax: +1 412 396 4196.
[3] Tel.: +1 573 882 1923; fax: +1 573 884 1151.

http://dx.doi.org/10.1016/j.rasd.2014.07.015
1750-9467/

## 1. Introduction

"Intelligence" has been more challenging to characterize in individuals with autism spectrum disorder (ASD) than in individuals with typical development (TD). ASD is a neurodevelopmental disorder characterized by impairments in social interactions and stereotyped and repetitive behaviors (American Psychiatric Association (APA), 2013). No IQ profile is representative of all individuals with ASD (Charman et al., 2011; Ghaziuddin & Mountain-Kimchi, 2004; Siegel, Minshew, & Goldstein, 1996; Williams, Goldstein, Kojkowski, & Minshew, 2008), and recent evidence suggests that the factor structure of intelligence and its relationship to adaptive functioning in this group is different from that found in individuals with TD (see Charman et al., 2011; Goldstein et al., 2008).

Measuring the intelligence of individuals with ASD is necessary for clinical assessment, educational service delivery, and research. The DSM-5 criteria for ASD include a specifier for cognitive abilities (APA, 2013), recognizing that affected individuals vary on this construct and that it may be important for identifying ASD subtypes (Grzadzinski, Huerta, & Lord, 2013). Establishment of an overall level of cognitive functioning is typically part of the process of qualifying a child for special education services (White, Scahill, Klin, Koenig, & Volkmar, 2007). Although not part of the diagnostic criteria, the related effect of ASD on cognitive functioning necessitates the consideration of the level of intelligence for most research in this area; the typical starting point for behavioral, neuroimaging, and neurofunctional studies is to individually- or group-match participants with ASD to controls on the basis of IQ, age, and gender or to use intellectual ability as a covariant in the statistical analyses. Important information about the cognitive and neurofunctional basis of ASD may be obscured if higher functioning individuals with ASD are compared to lower functioning individuals with TD (or vice versa). Therefore, the decision about what measure of intelligence to use in research is an important one (see Barbeau, Soulières, Dawson, Zeffiro, & Mottron, 2013; McGonigle-Chalmers & McSweeney, 2013).

Standardized measures of intelligence have generally been used to assess verbal individuals with ASD, including the various editions of the Stanford-Binet (e.g., SBV: Roid, 2003) and the Wechsler scales (Wechsler, 1981, 1991, 1999), a practice supported by some research (Filipek et al., 1999; Mottron, 2004). However, recent work has suggested that the Raven's Progressive Matrices (RPM; Raven, Raven, & Court, 1998) are better suited for capturing the nature of intelligence for individuals with ASD (Dawson, Soulières, Gernsbacher, & Mottron, 2007; Hayashi, Kato, Igarashi, & Kashima, 2008; Soulières, Dawson, Gernsbacher, & Mottron, 2011).

Dawson et al. (2007) were the first to suggest that the Wechsler scales may underestimate the intelligence of individuals with ASD and that the format of the RPM with its emphasis on fluid reasoning may be a more appropriate measure of intelligence for these individuals. They compared performance on the Wechsler assessments and the RPM in children and adults with and without ASD. Both the children and adults with ASD displayed dramatically higher RPM scores in comparison to age appropriate Wechsler assessments, while the TD groups exhibited no significant differences on these measures. The authors concluded that individuals with ASD are not impaired in fluid reasoning and suggested that conventional Wechsler IQ measures may underestimate the level of intelligence, which can lead to erroneous conclusions about the functioning level and cognitive profile of individuals with ASD (Dawson et al., 2007).

However, the IQ discrepancy (lower IQ abilities assessed in individuals with ASD in comparison to TD) across groups in the Dawson study makes a comparison of the differences on the RPM versus the Wechsler difficult to assess, as greater differences could occur in lower functioning individuals overall regardless of ASD diagnosis. That is, the RPM may be a more accurate assessment of nonverbal abilities for any group of individuals with low verbal skills. Indeed, results of a separate but related study suggest that the higher performance on the RPM as compared to the Wechsler measures primarily occurs for individuals with ASD with cognitive impairment (Bölte, Dziobek, & Poustka, 2009).

Soulières et al. (2011) also compared the performance of individuals with and without Asperger syndrome (ASP) (an ASD diagnosis characterized by relatively spared verbal abilities, lack of an early history of delayed language development, and IQ scores in the normal range, Klin, Volkmar, & Sparrow, 2000) on the RPM and the Wechsler scales to investigate whether the previously observed advantage on the RPM was also observed in more able individuals with ASD. Children and adults with and without ASP completed the RPM and the age appropriate Wechsler scale. Somewhat inconsistent with the findings by Dawson et al. (2007), both the adults with and without ASP had higher percentile performance on the RPM as compared to performance on the Wechsler IQ. However, this same pattern did not occur for the children, with both groups having similar performance on the RPM and Wechsler scales, though children with ASP had significantly higher RPM than Performance IQ scores.

Taken together, these previous studies suggest that individuals with ASD with low average and below average IQs obtain relatively higher scores on the RPM than on the Wechsler scales. The contention is, then, that the Wechsler scales are underestimating the cognitive abilities of individuals with ASD and leading to erroneous conclusions about their general intellectual abilities. However, the results for verbal, high-functioning children and adults with ASD (including those with ASP) have been mixed (Bölte et al., 2009; Chen, Planche, & Lemonnier, 2010; Dawson et al., 2007; Hayashi et al., 2008; Soulières et al., 2011; see Table 1), making it unclear how to proceed when selecting a measure for intelligence for these individuals. Given that measures of intelligence are used for establishing clinical diagnoses, for making educational decisions, and for group-matching in research studies of ASD, a continued investigation comparing the information provided by the RPM and the information provided by the Wechsler scales is warranted. Therefore, the objective of the current study was to further investigate the issue of intelligence measurement in individuals with ASD in a sample of verbal children and adults with high-functioning ASD and controls with TD to answer the question of to what extent, if any, their performance on the RPM differs from their performance on the Wechsler scales.

**Table 1**
Comparison of intelligence measures by research group and diagnostic group.

| Study | Sample | Units | ASD | | | | TD | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | RPM | FSIQ | VIQ | PIQ | RPM | FSIQ | VIQ | PIQ |
| Dawson et al. (2007) | Child | Percentiles | 56% | 26% | 26% | 31% | 72% | 70% | 70% | 67% |
| | Adult | Percentiles | 83.30% | 50.38% | – | – | 81.64% | 74.80% | – | – |
| Hayashi et al. (2008) | Child | Standard scores | 41.1[a] | 96.7 | 101.7 | 91.5 | 30.7[a] | 99.8 | 101.3 | 99.1 |
| Bölte et al. (2009) | Child/adult | Standard scores | 81.7 | 71.6 | 75.4 | 71.5 | 111.7 | 114.4 | 113.6 | 115 |
| | IQ > 85 | Approx. stand scores | ≈100 | ≈105 | ≈100 | ≈105 | ≈110 | ≈110 | ≈110 | ≈110 |
| | IQ < 85 | Approx. stand scores | ≈70 | ≈50 | ≈60 | ≈50 | ≈75 | ≈75 | ≈75 | ≈75 |
| Chen et al. (2010) | Child | RPM (mean rank) | 56.2 | 96.36[b] | – | – | 38.4 | "Normal Intelligence" | | |
| | IQ ≥ 90 | RPM (mean rank) | 63.7 | – | – | – | – | – | – | – |
| | IQ < 90 | RPM (mean rank) | 48.7 | – | – | – | – | – | – | – |
| Soulières et al. (2011) | Child | Percentiles | 59.05% | 52.12% | 63.74% | 41.08% | 71.83% | 69.26% | 69.78% | 64.78% |
| | Adult | Percentiles | 67.67% | 46.63% | 54.74% | 39.38% | 80.66% | 68.74% | 67.13% | 64.21% |
| Present Study | Child | Percentiles | 62.68% | 56.51% | 51.14% | 60.76% | 61.92% | 70.54% | 67.56% | 71.13% |
| | Adult | Percentiles | 25.84% | 40.65% | 34.23% | 47.48% | 45.03% | 62.17% | 62.46% | 60.89% |

*Note*: Version of the Raven's Progressive Matrices (RPM), Wechsler: full scale IQ (FSIQ), verbal IQ (VIQ), performance IQ (PIQ); means were not always reported in each study and missing data is indicated. Furthermore, an approximation of standard scores was made from a graph in Bölte et al. (2009).
[a] Total number correct responses.
[b] Standard score.

## 2. Methods

### 2.1. Participants

We had archived data that could be used to investigate this question comparing the difference in performance on the two measures of intelligence with respect to verbal, high-functioning individuals with ASD. This data consisted of sizeable numbers and well-matched groups of children and adults with and without ASD who had been administered both the RPM and a Wechsler IQ measure. A small number of these cases were excluded to achieve group matching on gender, age, and full scale and verbal IQ with the ASD group.

All participants with high functioning ASD and TD were recruited and assessed by the Subject Core of an autism research center at a major university as consecutive cases from the community meeting criteria for the research program. Two age groups consisting of 85 children between the ages of 8 and 16 years (37 ASD and 48 TD) and 66 adults between the ages of 17 and 46 years (31 ASD and 35 TD), were included in this study (see Tables 2 and 3).

The diagnosis for participants in the ASD groups was established through two structured research diagnostic instruments, the Autism Diagnostic Observation Schedule (ADOS: Lord et al., 2000) and the Autism Diagnostic Interview-Revised (ADI-R: Lord, Rutter, & Couteur, 1994), and expert clinical opinion. All participants in the ASD groups met cutoffs for the diagnosis of autism on these instruments. The ADI-R scores were not available for two of the children and three of the adults, but these individuals had reported early histories consistent with a diagnosis of idiopathic autism and met criteria for current presentation on the ADOS. The range of ADOS scores (communication + social interaction totals) for both the children and adults in the ASD groups were 10–24 with a mean of 15. A cut-off score of 10 is consistent with a diagnosis of *Autism* on this instrument. All participants in the ASD groups had FSIQs greater or equal to 70, were able to communicate in complete

**Table 2**
Descriptive statistics, unstandardized and standardized effect sizes and confidence intervals for children with and without ASD.

| | ASD $n = 37$ $M$ (SD) | TD $n = 48$ $M$ (SD) | $\chi^2$ ($p$) | $M_{diff}$ [95% CI] | $d_{unb}$ [95% CI] |
|---|---|---|---|---|---|
| Age (in years) | 11.85 (2.35) | 12.09 (2.27) | – | 0.24 [−0.76, 1.24] | 0.10 [−0.34, 0.54] |
| VIQ | 107.08 (17.00) | 110.38 (10.92) | – | 3.30 [−2.74, 9.34] | 0.24 [−0.21, 0.68] |
| PIQ | 101.08 (17.07) | 108.44 (10.97) | – | 7.36 [1.29, 13.43] | 0.52 [0.07, 0.97] |
| FSIQ | 104.46 (16.70) | 110.19 (11.28) | – | 5.73 [−0.32, 11.78] | 0.41 [−0.04, 0.85] |
| RPM ($t$-score) | 41.16 (10.49) | 41.42 (7.04) | – | 0.26 [−3.53, 4.05] | 0.03 [−0.41, 0.47] |
| SES[a] | – | – | 1.45 (0.23) | – | – |
| Education | – | – | 0.88 (0.35) | – | – |
| Race | – | – | 0.77 (0.38) | – | – |
| Gender (male:female) | 35:2 | 43:5 | 0.69 (0.41) | – | – |

*Note*: $\chi^2$ values derived from Cochran–Mantel–Haenszel tests. Effect size estimates and confidence intervals assumed independence. All Cohen's $d$ values were calculated using the pooled within-groups standard deviation as the standardizer. $M_{diff}$ – unstandardized mean difference; CI – confidence interval; $d_{unb}$ – unbiased estimate of Cohen's $d$. Gender variable indicates the number of male to female participants within each group.
[a] ASD: $n = 35$, TD: $n = 42$.

**Table 3**
Descriptive statistics and unstandardized and standardized effect sizes and confidence intervals for adults with and without ASD.

| | ASD<br>$n=31$<br>*M* (SD) | TD<br>$n=35$<br>*M* (SD) | $\chi^2$ (*p*) | $M_{diff}$ [95% CI] | $d_{unb}$ [95% CI] |
|---|---|---|---|---|---|
| Age (in years) | 26.78 (8.51) | 27.64 (9.07) | – | 0.86 [−3.48, 5.20] | 0.10 [−0.40, 0.60] |
| VIQ | 99.19 (15.32) | 105.54 (12.81) | – | 6.35 [−0.57, 13.27] | 0.45 [−0.06, 0.95] |
| PIQ | 92.13 (12.01) | 106.00 (12.90) | – | 13.87 [7.72, 20.02] | 1.10 [0.56, 1.64] |
| FSIQ | 95.65 (13.69) | 106.29 (13.61) | – | 10.64 [3.92, 17.36] | 0.77 [0.25, 1.29] |
| RPM (*t*-scores) | 42.13 (9.62) | 48.40 (7.26) | – | 6.27 [2.11, 10.43] | 0.73 [0.22, 1.25] |
| SES[a] | – | – | 0.39 (.53) | – | – |
| Education[b] | – | – | 3.72 (.05) | – | – |
| Race | – | – | 1.13 (.29) | – | – |
| Gender (male:female) | 25:6 | 31:4 | 0.79 (.37) | – | – |

*Note*: $\chi^2$ values derived from Cochran–Mantel–Haenszel tests. Effect size estimates and confidence intervals assumed independence. All Cohen's *d* values were calculated using the pooled within-groups standard deviation as the standardizer. $M_{diff}$ – unstandardized mean difference; CI – confidence interval; $d_{unb}$ – unbiased estimate of Cohen's *d*. Gender variable indicates the number of male to female participants within each group.
[a] (ASD: $n=23$, TD: $n=32$).
[b] (ASD: $n=29$, TD: $n=35$).

spoken sentences, did not have attention or behavioral problems that prevented them from completing testing, did not have any associated or causative genetic, metabolic, or infectious conditions, were in good medical health, and had no history of seizures, birth injury, or head trauma.

For purposes of comparison of the relationship of performance on the Wechsler measures and RPM, TD participants were recruited from the community with the same SES as the family of origin of the participants with ASD. Potential TD participants were screened by telephone questionnaires, interviews, and psychometric evaluations. TD participants were excluded if found to have a family history of ASD, developmental cognitive disorder, learning disability, affective disorder, anxiety disorder, schizophrenia, obsessive-compulsive disorder, or other neurologic or psychiatric disorders thought to have a genetic component.

The ASD and TD groups were matched by age, gender, SES, education, race, and VIQ. The group-matching on the Wechsler VIQ scores was used to equate the general performance level of the groups on a measure that was thought to be somewhat conceptually independent from the skills being measured by the RPM. When matched on VIQ, the mean PIQ scores for the ASD groups were significantly lower than the TD groups. A lower PIQ than VIQ in the ASD group is consistent with the pattern reported in Soulières et al. (2011) comparison of Wechsler FSIQ and RPM performance in a group of children and adults with ASP. In that study, both the children and adult groups with ASP had a significant advantage of RPM over Wechsler PIQ scores. Therefore, the general Wechsler IQ profile of our participants with ASD was consistent with previously reported research in this area. More importantly, if nonverbal performance is more accurately represented by the RPM than by the Wechsler PIQ, the lower mean PIQ scores of our participants with ASD potentially created a bias in favor of finding an RPM advantage.

Approval for the study was granted by the Institutional Review Board at a major university. Written informed consent was obtained from participants and/or guardians prior to testing.

### 2.2. Measures

All tests were completed as part of a standard neuropsychological test battery and administered by trained examiners at an autism research center associated with a major university in the standard manner. All participants completed the standard version of the RPM in which a matrix of complex patterns or designs with a piece missing is presented and the participant has to select the missing piece from an array of possible options (Raven et al., 1998). The test is comprised of 60 problems divided into five sets of 12 problems each. The items progress in difficulty from more simple to increasingly complex relationships with later items building on earlier established relationships. The RPM was designed to be used with both children and adults. The RPM was administered individually and was self-paced with no imposed time limits. RPM raw scores were converted to percentiles using the age appropriate United State Norms. Child participants were administered the Wechsler Intelligence Scale for Children, third edition (WISC-III: Wechsler, 1991). Adult participants were administered the Wechsler Adult Intelligence Scale, Revised (WAIS-R: Wechsler, 1981). Percentile ranks for the Wechsler scores were used for the data analyses to be consistent with earlier reported work in this area (i.e., Dawson et al., 2007; Soulières et al., 2011).

### 2.3. Analytic strategy

The primary question for our analyses was one of magnitude (e.g., "To what extent do two means differ?") with respect to the performance by each age/diagnostic group on the RPM and Wechsler scales. The American Psychological Association has suggested that, whenever possible, effect size point estimates and confidence intervals are preferable to null hypothesis significance testing (NHST), particularly when investigating psychological constructs that are best conceptualized along a continuum (American Psychological Association, 2010). Thinking in terms of estimation and precision, that is in terms of effect size estimates and confidence intervals [CIs]), is preferable to NHST thinking for a multitude of reasons and is the approach adopted here. For example, CIs provide rather useful information about expected mean differences under replication; *p* values, by contrast, can take on virtually any value upon replication and thus provide very little information about what is expected in future studies. In short, much more information can be gleaned from effect size point estimates and CIs than from *p* values (for a review, see Cumming, 2012). Therefore, the current report provides unstandardized and standardized effect size estimates and corresponding CIs for all independent and paired comparisons. Cohen's *d* effect size estimates for paired means were calculated using the square root of the average variances as the standardizer; Cohen's *d* effect size estimates for independent means were calculated using the pooled within-groups standard deviation as the standardizer. All 95% CIs for *d* were calculated using approximations of the noncentral *t* distribution (see Algina & Keselman, 2003; Cumming, 2012; Cumming & Fidler, 2009; Rosnow & Rosenthal, 2009).

### 2.4. Power analysis

In addition to facilitating dichotomous thinking, recent simulations have indicated that NHST can promote an unsettling number of false positive publications (see Bakker, van Dijk, & Wicherts, 2012). To directly address this issue, a power analysis was conducted to determine whether the sample sizes used in the current study would have been adequately powered to detect IQ measure differences if NHST had been employed as a conduit for statistical inference. Power calculations were based on the reported estimated effect sizes of RPM and FSIQ differences among individuals with ASD reported by Dawson et al. (2007), as their study most directly replicated the crux of the primary tests reported here. Dawson et al. yielded paired Cohen's *d* values ranging from 0.78 to 1.29. The G*Power statistical program (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that the current sample of 37 children and 31 adults with autism would have had excellent power ($>$0.99 in both groups) to detect mean differences using a two-tailed, paired *t* test with $\alpha = .05$. In other words, the current study was extremely well-powered to detect within-group comparisons based on the effect size estimates reported in previous research. Even so, *p* values are reported sparingly for the reasons mentioned above, except in the case of $\chi^2$ or multiple degree of freedom tests, but enough information is provided in the sections that follow to directly and easily calculate *p* values. It should be noted that the pattern of results reported here is consistent with the results from traditional NHSTs.

## 3. Results

### 3.1. Child samples

The effect size point estimates and CIs for the between-group IQ test comparisons may be seen in Table 2. In general, mean differences ranged from extremely small (RPM) to moderate (PIQ), suggesting that the samples were generally well-matched.

The primary focus of the current report was to determine the extent to which the RPM performance was higher or lower than the Wechsler performance as measured by the PIQ, VIQ, and FSIQ among children with ASD and TD. Fig. 1 depicts the mean percentile scores of all IQ measures by group; Table 4 shows the percentile IQ means and standard deviations by group. Focusing on the children with ASD first, the largest differences were between RPM and PIQ scores, $M_{diff} = 11.54$ [1.34, 21.74], $d_{unb} = 0.35$ [0.04, 0.68], and between the VIQ and PIQ scores, $M_{diff} = 9.63$ [1.72, 17.54], $d_{unb} = 0.31$ [0.05, 0.57], as children with ASD had a higher RPM compared to PIQ performance. Because the RPM and VIQ scores were of similar magnitude, $M_{diff} = 1.91$ [−9.31, 13.14], $d_{unb} = 0.06$ [−0.29, 0.41], the difference between RPM and FSIQ scores was slightly mitigated, $M_{diff} = 6.16$ [−4.12, 16.45], $d_{unb} = 0.19$ [−0.12, 0.50], but the RPM score was higher than the Wechsler score. The pattern of effects was slightly different among children with TD. Specifically, the RPM was moderately lower than VIQ scores, $M_{diff} = -9.21$

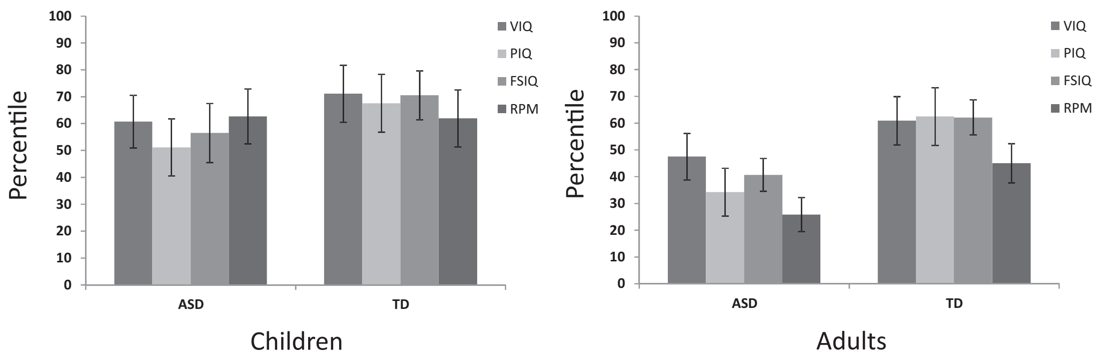


**Fig. 1.** Mean percentile performance of children (left) and adults (right) with ASD and with typical development (TD) on the age appropriate version of the Wechsler Intelligence Scales and the Raven's Progressive Matrices. FSIQ – full-scale IQ. VIQ – verbal IQ. PIQ – performance IQ. Error bars denote 95% confidence intervals for the mean.

**Table 4**
Mean percentile scores and standard deviations by IQ test and developmental group.

| | Children | | Adults | |
|---|---|---|---|---|
| | ASD | TD | ASD | TD |
| | *n* = 37 | *n* = 48 | *n* = 31 | *n* = 35 |
| | *M* (SD) | *M* (SD) | *M* (SD) | *M* (SD) |
| VIQ | 60.76 (29.5) | 71.13 (21.13) | 47.48 (29.95) | 60.89 (25.28) |
| PIQ | 51.14 (31.81) | 67.56 (22.63) | 34.23 (24.78) | 62.46 (26.3) |
| FSIQ | 56.51 (31.83) | 70.54 (21.99) | 40.65 (27.78) | 62.17 (26.15) |
| RPM | 62.68 (32.33) | 61.92 (25.23) | 25.84 (28.83) | 45.03 (31.51) |

[−16.02, −2.4], $d_{unb} = -0.39$ [−0.69, −0.10], and FSIQ scores, $M_{diff} = -8.63$ [−15.63, −1.62], $d_{unb} = -0.36$ [−0.66, −0.06], in this group. Taken together, while the differences between RPM and FSIQ scores were of similar magnitude across developmental groups, the effects were in opposing directions. That is, RPM percentile scores were slightly *higher* than FSIQ scores among children with ASD but moderately *lower* FSIQ scores among children with TD. An alternative way of expressing this relationship is that children with ASD had slightly *lower* Wechsler scales FSIQ scores relative to their RPM percentile scores. However, the Wechsler scales FSIQ scores were *higher* relative to RPM performance in children with TD.

### *3.2. Adult samples*

The effect sizes and CIs for between-group differences were slightly larger in the adult sample (see Table 3). For example, the estimated standardized effect size for PIQ scores among adults was approximately twice as large as the effect for children. For all intents and purposes, however, the adult sample was well-matched on age but less well-matched than the child sample on the IQ subtests.

The most striking finding is that the pattern of IQ percentile score means for the adults with ASD differed from the pattern of results for the children with ASD (see Fig. 1 and Table 4). Specifically, not only was the RPM *lower* than FSIQ scores in adults with ASD, but the estimated magnitude of this difference was moderate in size, $M_{diff} = -14.81$ [−23.78, −5.83], $d_{unb} = -0.51$ [−0.85, −0.19]. The RPM-FSIQ difference was primarily driven by a larger difference between RPM and VIQ scores, $M_{diff} = -21.65$ [−32.3, −10.99], $d_{unb} = -0.72$ [−1.13, −0.34] than by RPM and PIQ scores, $M_{diff} = -8.39$ [−17.73, 0.96], $d_{unb} = -0.30$ [−0.65, 0.03]. Adults with TD had the same pattern of means as children with TD, though the extent to which the RPM was lower than FSIQ scores was slightly larger, $M_{diff} = -17.14$ [−25.32, −8.97], $d_{unb} = -0.58$ [−0.89, −0.28].

## 4. Discussion

The present study compared the performance on the RPM and Wechsler Scales in higher functioning, verbally able children and adults with and without ASD to further address the issue of the differences in characterization of intelligence in ASD based on which tool was selected to make the measurement. Consistent with the findings of Soulières et al. (2011), in the current study, children with ASD had higher RPM percentiles in comparison to their PIQ percentile. However, a different pattern occurred with respect to the RPM and Wechsler FSIQ relationship. In the Soulières et al. study no difference occurred for the RPM and Wechsler FSIQ measures for the children with ASP. In the current study, children with ASD displayed higher RPM percentiles compared to FSIQ percentiles as measured by the WISC-III. A different pattern of results emerged for the adults with ASD. Soulières et al. reported lower WAIS percentiles compared to RPM percentiles in adults with ASD. However, in the current study, adults with ASD had lower RPM percentiles compared to WAIS FSIQ percentiles. Furthermore, that pattern for the relationship between the RPM and FSIQ was also obtained for children and adults with TD.

Considered together, the results of the various studies suggest that the RPM may produce relatively higher scores for both low- and high-verbal children with ASD or ASP than the Wechsler PIQ measure. However, the RPM demonstrated no particular advantage for high-verbal children with ASD or ASP as compared to Wechsler FSIQ measures. The standardized 95% CI for PIQ scores among children suggests that there is an 83% chance that an exact replication of the current study could yield mean differences as small as 0.07 or as large as 0.97. The CIs also suggest that in many cases, although comparatively less likely, children with ASD could score higher on the various IQ tests than typically developing children in the population of these two groups upon replication. Therefore, the RPM may be an appropriate measurement tool when assessing nonverbal cognitive ability in both low- and high-verbal children with ASD, but the current findings suggest that it confers no particular advantage when used as an alternative to the Wechsler scales for verbal children with ASD.

Across studies, the results for the adult groups are less consistent than for the child groups. Why the various adult groups performed differently on the RPM is not clear. The aspect of cognition that is purported to be measured by the RPM, a specific kind of "meaning making ability" (Raven, 2008), may actually differ in these groups. Alternately, differences in the level of verbal ability, cognitive functioning, and experiential knowledge may have influenced the results. For example, as previously noted, groups were intentionally matched on a variable thought to be independent of RPM performance (VIQ) and findings may be dependent on this matching strategy.[5] The important point is that no consistent pattern is emerging when the RPM is used with older, verbal individuals with ASD, suggesting great care must be taken when selecting intelligence measures for adults with ASD.

The implications of the current study when considered with the results of Dawson et al. (2007), Soulières et al. (2011) study, and studies of similar standardized assessments in clinical samples (Silverman et al., 2010) are that the cognitive differences characteristic of ASD cannot be reduced to or fully revealed by performance on an IQ test despite test form. This is partially due to the limitations inherent in the use of standardized tests to measure intelligence in general. However, the nature of the cognitive differences in ASD creates a particular challenge for the use of formal or standardized tests in this manner.

Previous research has indicated that the cognitive skills of children and adults with ASD are characterized by differences in processing across the cognitive domains, particularly during tasks that are challenging, due to the nature of information or the constraints of testing, and differences are only revealed by using a large battery of tests that provide measurement across cognitive domains (Minshew, Goldstein, & Siegel, 1997; Williams, Goldstein, & Minshew, 2006). Based on this research, the cognitive profile of individuals with ASD is not captured by the presence or absence of certain specific cognitive skills. Rather, cognitive tasks become more challenging for individuals with ASD at levels lower than would be predicted by their overall general cognitive ability, an effect that is particularly seen as the demand for integration of information increases or as larger amounts of information without an inherent structure must be processed. The 'complexity' in this model relates to the dynamics of the information processing, not a particular type of information per se. That is, individuals with ASD are likely using different underlying neural responses that are susceptible to disruption (Just, Keller, Malave, Kana, & Varma, 2012; Minshew, Williams, & McFadden, 2008) but when this disruption will occur may be individually or situationally specific. Measures of intelligence provide information about the general level of cognitive resources but do not present a complete picture of cognitive processing capabilities in ASD. Understanding the cognitive challenges of ASD is most apparent in real world situations where the flow of information is dynamic and the demands for integration and decision-making are temporaneously brief.

In other words, individuals with ASD process information differently and, therefore, performance on an IQ test of any kind may not be predictive of their ability to function in the same way as individuals with TD on whom IQ tests are normed. Differential performance may occur on the RPM because aspects of the test may be adaptive for some individuals with ASD allowing them to perform better on this measure, such as for individuals with ASD who have less developed or less consistent verbal abilities. By the same token, differences may not be as adaptive for other individuals with ASD resulting in equivocal performance as compared to another standardized measure of cognitive functioning.

Another implication of the current study is relevant to the use of intelligence test scores for the purposes of research or clinical practice. The points made by Dawson et al. (2007) and Soulières et al. (2011) regarding the underestimation of the cognitive abilities of individuals with ASD should be of concern when choosing any instrument to use with individuals with ASD and when interpreting the results of their performance on that instrument. The RPM would appear to provide an appropriate measure of "meaning making" abilities in individuals with ASD, particularly for children or for those who are less verbally able. However, RPM performance was lower than Wechsler performance in the verbally able adults with ASD in the current study, suggesting that multiple elements should be considered (e.g. context, situation, abilities assessed, methods used by similar studies) before selecting a measure of cognitive functioning in verbal individuals with ASD.

The current study is not all encompassing, and future research should be undertaken to further shed light on this complex issue. Specifically, it remains unclear what role gender may play in these results. In typically developing individuals, studies investigating aspects of intelligence and gender report that males and females display different patterns of strengths and weakness in comparison to each other (Halpern, 1997; Lynn & Irwing, 2004; Lynn, Raine, Venables, Mednick, & Irwing, 2005). Previous research investigating potential sex differences in individuals with ASD has been mixed, with some studies reporting that males with ASD have higher IQs than females with ASD (Lord & Schopler, 1985; Volkmar, Szatmari, & Sparrow, 1993) and other more recent studies reporting equivalent performance (Mandy et al., 2012). In the present study, the number of female participants is too small to make more conclusive interpretations about the role of gender.[6] Given the inconclusive literature on this topic in individuals with ASD, future research investigating the role of gender on this topic is warranted.

Adequate caution should be exercised in interpreting the results of these measures, especially when using them to predict academic or adaptive functioning. Neither the Wechsler scales nor the RPM fully capture 'intelligence' or cognitive functioning in ASD with respect to either strengths or weaknesses. The nature of intelligence is multifaceted, therefore, multiple lines of inquiry as to its nature, strengths, and challenges in any individual are required whether or not that individual is on the autism spectrum.

[5] Interestingly, when participants were matched on PIQ instead of VIQ and analyses were re-run, our results did not change dramatically.

[6] Analyses resulted in no statistically significant differences between males and females on the RPM among children or adults ($p > 0.05$). Furthermore, all effect sizes were re-calculated without female participants, and the magnitude of the point estimates did not change in any appreciable way (the largest standardized effect size change across all analyses was 0.12; the average change was 0.08). Therefore, female participants were retained in all reported analyses.

## Acknowledgements

We would like to thank participants and families for their contribution to this study, and Dana Barvinchak, Kristin Weisberg, Natasha Reddy, and John J. Markiewicz for their assistance in data collection. We acknowledge the support of the grant sponsor: National Institute of Neurological Disorders and Stroke (NINDS), grant number: NS33355; grant sponsor: National Institute of Child Health and Human Development (NICHD), grant number: HD35469, a Collaborative Program of Excellence in Autism (CPEA); grant sponsor: National Institute on Deafness and other Communication Disorders (NIDCD), grant number: K23DC006691 to Dr. Williams; and, grant sponsor: Edith L. Trees Charitable Trust funding without which this work would not have been possible.

## References

Algina, J., & Keselman, H. J. (2003). Approximate confidence intervals for effect sizes. *Educational and Psychological Measurement, 63*, 537–553.
American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*, 543–554.
Barbeau, E. B., Soulières, I., Dawson, M., Zeffiro, T. A., & Mottron, L. (2013). The level and nature of autistic intelligence III: Inspection time. *Journal of Abnormal Psychology, 122*, 295–301.
Bölte, S., Dziobek, I., & Poustka, F. (2009). Brief report: The level and nature of autistic intelligence revisited. *Journal of Autism and Developmental Disorders, 39*(4), 678–682.
Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: Data from the Special Needs and Autism Project (SNAP). *Psychological Medicine, 41*, 619–627.
Chen, F., Planche, P., & Lemonnier, E. (2010). Superior nonverbal intelligence in children with high-functioning autism or Asperger's syndrome. *Research in Autism Spectrum Disorders, 4*(3), 457–460.
Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.
Cumming, G., & Fidler, F. (2009). Confidence intervals: Better answers to better questions. *Zeitschrift fuer Psychologie,* Journal of Psychology*217*, 15–26.
Dawson, M., Soulières, I., Gernsbacher, M. A., & Mottron, L. (2007). The level and nature of autistic intelligence. *Psychological Science, 18*(8), 657–662.
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.
Filipek, P. A., Accardo, P. J., Baranek, G. T., Cook, E. H., Dawson, G., Gordon, B., et al. (1999). The screening and diagnosis of autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 29*(6), 439–484.
Ghaziuddin, M., & Mountain-Kimchi, K. (2004). Defining the intellectual profile of Asperger syndrome: Comparison with high-functioning autism. *Journal of Autism and Developmental Disorders, 34*(3), 279–284.
Goldstein, G., Allen, D., Minshew, N., Williams, D., Volkmar, F., Klin, A., et al. (2008). The structure of intelligence in children and adults with high functioning autism. *Neuropsychology, 22*(3), 301–312.
Grzadzinski, R., Huerta, M., & Lord, C. (2013). DSM-5 and autism spectrum disorders (ASDs): An opportunity for identifying ASD subtypes. *Molecular Autism, 4*, 12.
Halpern, D. F. (1997). Sex differences in intelligence. *American Psychologist, 52*, 1091–1102.
Hayashi, M., Kato, M., Igarashi, K., & Kashima, H. (2008). Superior fluid intelligence in children with Asperger's disorder. *Brain and Cognition, 66*(3), 306–310.
Just, M. A., Keller, T. A., Malave, V. L., Kana, R. K., & Varma, S. (2012). Autism as a neural systems disorder: A theory of frontal-posterior underconnectivity. *Neuroscience & Biobehavioral Reviews, 36*(4), 1292–1313.
Klin, A., Volkmar, F. R., & Sparrow, S. S. (2000). *Asperger syndrome*. New York, NY: The Guilford Press.
Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., 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*(3), 205–223.
Lord, C., Rutter, M., & Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*(5), 659–685.
Lord, C., & Schopler, E. (1985). Brief report: Differences in sex ratios in autism as a function of measured intelligence. *Journal of Autism and Developmental Disorders, 15*(2), 185–193.
Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*, 481–498.
Lynn, R., Raine, A., Venables, P. H., Mednick, S. A., & Irwing, P. (2005). Sex differences on the WISC-R in Mauritius. *Intelligence, 33*, 527–533.
Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2012). Sex differences in autism spectrum disorder: Evidence from a large sample of children and adolescents. *Journal of Autism and Developmental Disorders, 42*, 1304–1313.
McGonigle-Chalmers, M., & McSweeney, M. (2013). The role of timing in testing nonverbal IQ in children with ASD. *Journal of Autism and Developmental Disorders, 43*, 80–90.
Minshew, N., Goldstein, G., & Siegel, D. (1997). Neuropsychologic functioning in autism: Profile of a complex information processing disorder. *Journal of the International Neuropsychological Society, 3*(4), 303–316.
Minshew, N. J., Williams, D. L., & McFadden, K. (2008). Information processing, neural connectivity and neuronal organization. In A. Zimmerman (Ed.), *Autism: Current theories and evidence* (pp. 381–405). Totowa, NJ: Humana Press.
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*(1), 19–27.
Raven, J. (2008). The Raven progressive matrices tests: Their theoretical basis and measurement model. In J. Raven & J. Raven (Eds.), *Uses and abuses of intelligence: Studies advancing Spearman and Raven's quest for non-arbitrary metrics* (pp. 17–68). Unionville, NY: Royal Fireworks Press.
Raven, J., Raven, J., & Court, J. (1998). *Raven manual: Section 3. The standard progressive matrices*. San Antonio, TX: Harcourt Assessment Inc.
Roid, G. (2003). *Stanford-Binet intelligence scales* (5th ed.). Chicago, IL: Riverside.
Rosnow, R., & Rosenthal, R. (2009). Effect sizes: Why, when and how to use them. *Zeitschrift fuer Psychologie,* Journal of Psychology*217*, 6–14.
Siegel, D., Minshew, N., & Goldstein, G. (1996). Wechsler IQ profiles in diagnosis of high-functioning autism. *Journal of Autism and Developmental Disorders, 26*(4), 389–406.
Silverman, W., Miezejeski, C., Ryan, R., Zigman, W., Krinsky-McHale, & Urv, T. (2010). Stanford-Binet and WAIS IQ differences and their implications for adults with intellectual disability (aka mental retardation). *Intelligence, 38*, 242–248.
Soulières, I., Dawson, M., Gernsbacher, M. A., & Mottron, L. (2011). The level and nature of autistic intelligence II: What about Asperger syndrome? *PLOS ONE, 6*, e25372.
Volkmar, F. R., Szatmari, P., & Sparrow, S. S. (1993). Sex differences in pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 23*, 579–591.
Wechsler, D. (1981). *Wechsler adult intelligence scale-revised*. New York, NY: Psychological Corporation.
Wechsler, D. (1991). *Wechsler intelligence scale for children* (3rd ed.). San Antonio, TX: Psychological Corporation.
Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. San Antonio, TX: Psychological Corporation.

White, S. W., Scahill, L., Klin, A., Koenig, K., & Volkmar, F. R. (2007). Educational placements and service use patterns of individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*, 1403–1412.

Williams, D., Goldstein, G., & Minshew, N. (2006). Neuropsychologic functioning in children with autism: Further evidence of disordered complex information processing. *Child Neuropsychology, 12*, 279–298.

Williams, D. L., Goldstein, G., Kojkowski, N., & Minshew, N. J. (2008). Do individuals with high functioning autism have the IQ profile associated with nonverbal learning disability? *Research in Autism Spectrum Disorders, 2*(2), 353–361.