Contemporary School Psychology (2019) 23:115–125
https://doi.org/10.1007/s40688-018-0172-3

Diagnostic Utility of the WISC-IV GAI > CPI Cognitive Score Profile for a Referred Sample of Children and Adolescents with Autism

Kara M. Styck[1] · Miriam S. Aman[2] · Marley W. Watkins[3]

Abstract

Individuals with autism spectrum disorder (ASD) are hypothesized to exhibit relative strengths in verbal and non-verbal reasoning and weaknesses in working memory and speed of information processing. The purpose of the present investigation was to determine the degree to which this cognitive profile as measured by the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV; Wechsler 2003a) cognitive proficiency index (CPI; measure of working memory and processing speed) and general ability index (GAI; measure of verbal and non-verbal reasoning) could accurately distinguish between a referred sample of 79 school-aged students diagnosed with ASD and two non-clinical comparison groups: (a) 2200 children in the WISC-IV standardization sample and (b) 216 school-aged students referred for psychoeducational testing whose school-based evaluations did not result in a diagnosis. Results indicated that the ASD sample exhibited significantly lower mean scores on the CPI when compared to the two control groups. However, diagnostic utility statistics indicated that a randomly selected participant from the ASD subgroup would exhibit a larger difference between the GAI and CPI than a randomly selected participant from the two control groups 51.9–66.0% of the time. Consequently, the GAI > CPI cognitive score profile exhibits low diagnostic accuracy for individuals with ASD. Psychologists who work in applied settings are cautioned against using group trends to guide decision-making for individual clients.

Keywords Diagnostic utility · ROC · IQ · Autism

Applied Utility of the WISC-IV GAI > CPI Cognitive Score Profile for Children and Adolescents with Autism

The cognitive score profiles of individuals with autism spectrum disorder[1] (ASD) have been studied extensively with over 30 peer-reviewed journal articles published to date (e.g., see Barnhill et al. (2000), Happe (1994), and Zander and Dahlgren (2010) for more comprehensive historical reviews). Cognitive score profiles have been touted as useful for diagnosing ASD (e.g., Lincoln et al. 1988; Mayes and Calhoun 2003), differentiating ASD from other pervasive developmental disorders (e.g., Foley-Nicpon et al. 2012; Koyama et al. 2007; Planche and Lemonnier 2012), and informing treatment decisions (e.g., Mayes and Calhoun 2008).

The most commonly cited cognitive score profile for individuals with ASD is one in which the shape of the profile is marked by relative strengths in verbal and non-verbal reasoning and weaknesses in working memory and speed of information processing (Calhoun and Mayes 2005; Foley-Nicpon et al. 2012; Mayes and Calhoun 2003, 2008; Oliveras-Rentas et al. 2012; Zander and Dahlgren 2010). Evidence in support of the presence and utility of this profile is largely drawn from the nomothetic perspective that, as a group, individuals diagnosed with ASD tend to have higher subtest and index scores on tasks that measure the former cognitive domains relative to subtest and index scores on tasks that measure those latter

---

[1] Use of the term autism spectrum disorder (ASD) herein refers to the current definition in the *Diagnostic and Statistical Manual of Mental Disorders—Fifth Edition* (2013) that includes other conditions formerly referred to as pervasive developmental disorders, such as Asperger's disorder and pervasive developmental disorder-not otherwise specified, unless otherwise noted.

✉ Kara M. Styck
Kara.Styck@utsa.edu

[1] Department of Educational Psychology, University of Texas at San Antonio, 501 W. Cesar E. Chavez Blvd, San Antonio, TX 78207, USA

[2] Dallas Independent School District, Dallas, TX, USA

[3] Department of Educational Psychology, Baylor University, Waco, TX, USA

cognitive domains. This "group profile" for individuals with ASD has been most widely documented on the Wechsler scales, which is unsurprising given that they are the most common individually administered intelligence tests in applied settings (Braden 2013). On contemporary versions of the Wechsler scales, individuals with ASD have specifically exhibited higher mean scores on the verbal comprehension index (VCI) and perceptual reasoning index (PRI) compared to the working memory index (WMI) and processing speed index (PSI; Foley-Nicpon et al. 2012; Mayes and Calhoun 2008; Oliveras-Rentas et al. 2012).

Mayes and Calhoun (2008) evaluated the cognitive score profiles of 54 children and adolescents aged 6–14 years old diagnosed with high-functioning autism who were administered the Wechsler Intelligence Scale for Children—Fourth edition (WISC-IV; Wechsler 2003a). The researchers defined "high functioning" as a full-scale intelligence quotient (FSIQ) score ≥ 70. Results indicated that mean VCI and PRI scores were significantly higher than mean WMI and PSI scores. Mayes and Calhoun also reported that mean VCI and PRI scores were significantly higher than the normative mean, whereas mean WMI and PSI scores were significantly lower than the normative mean. At the subtest level, the Coding, Symbol Search, Letter-Number Sequencing, and Digit Span subtests all exhibited the lowest mean scores for the sample, and each was significantly lower than the normative mean. Mean scores on Vocabulary and Similarities were significantly higher than mean scores on Comprehension, and mean scores on Picture Concepts and Matrix Reasoning were significantly higher than mean scores on Block Design. At the individual level, Mayes and Calhoun (2008) reported that the WMI and PSI "…were lower than *or equal* [emphasis added] to VCI and PRI in 74% of the children" (pp. 431–432) in their sample leaving the question unanswered as to how many participants displayed the hypothesized profile consisting of WMI and PSI scores that were *lower* than scores on the VCI and PRI.

Foley-Nicpon et al. (2012) reported a similar pattern of results in an independent sample of 42 "academically and/or cognitively gifted" (p. 79) school-aged students with ASD further disaggregated into two subgroups: high-functioning autism ($n = 20$) and Asperger's disorder ($n = 22$). Results indicated that mean VCI and PRI scores were higher than mean WMI and PSI scores across the two participant subgroups replicating the pattern of results reported in Mayes and Calhoun (2008). However, no statistical tests were performed to determine the significance of these observed differences. Foley-Nicpon et al. noted that mean VCI scores for participants in the Asperger's disorder subgroup were significantly higher than mean VCI scores for participants in the high-functioning autism subgroup, and mean PSI scores for participants in the high-functioning autism subgroup were significantly higher than mean PSI scores for participants in the Asperger's disorder subgroup. At the subtest level, participants in the Asperger's disorder subgroup displayed significantly higher mean scores on Vocabulary compared to the high-functioning autism subgroup. The reverse pattern was noted for performance on the Symbol Search subtest with participants in the high-functioning autism subgroup exhibiting significantly higher mean scores when compared to the Asperger's disorder subgroup. At the individual level, Foley-Nicpon et al. noted that 28% of participants in the high-functioning autism subgroup exhibited PRI > VCI, 56% exhibited PRI = VCI, and 16% exhibited VCI > PRI, whereas 15% of the participants in the Asperger's disorder subgroup exhibited PRI > VCI, 60% exhibited PRI = VCI, and 25% exhibited VCI > PRI. However, no individual statistics were reported for the frequency of participants who displayed VCI and PRI scores that were greater than WMI and PSI scores for either participant subgroup.

The general pattern of results reported in Mayes and Calhoun (2008) and Foley-Nicpon et al. (2012) were partially replicated in a third study by Oliveras-Rentas et al. (2012) in a sample of 56 children and adolescents diagnosed with autism, Asperger's disorder, or pervasive developmental disorder-not otherwise specified. Results indicated that mean PSI scores for the aggregated sample were significantly lower than the normative mean, albeit significant mean differences did not emerge for the VCI, PRI, or WMI. At the subtest level, mean scores on Coding, Symbol Search, and Comprehension were significantly lower than the normative mean, and mean scores on Similarities and Matrix Reasoning were significantly higher than the normative mean. Mean within-group differences were also reported for WISC-IV index and subtest scores. Results indicated that mean scores on the PSI were significantly lower than mean scores on the PRI, VCI, and WMI, whereas mean scores on the PRI were significantly higher than mean scores on the WMI and VCI. At the subtest level, mean scores on Comprehension and Symbol Search were significantly lower than mean scores on Block Design, Similarities, Vocabulary, Digit Span, and Matrix Reasoning. No individual statistics were reported.

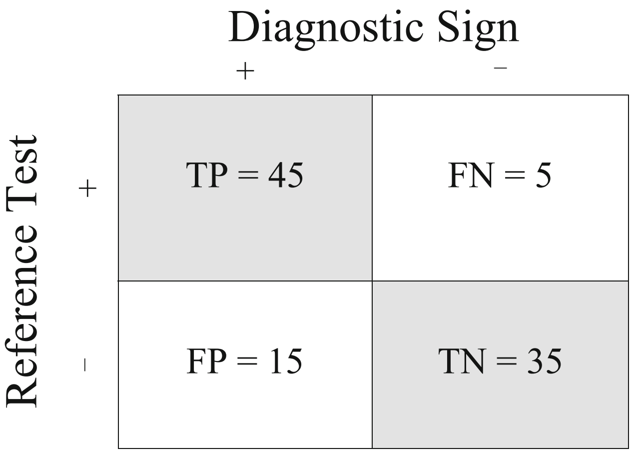
Nevertheless, these results led Mayes and Calhoun (2008) to conclude "the relative strengths in visual and verbal reasoning revealed on the WISC-IV for children with high-functioning autism have educational implications (i.e., teaching to the child's visual and verbal strengths while compensating for the writing, attention, processing speed, language comprehension, and social reasoning weaknesses)" (p. 434). Oliveras-Rentas et al. (2012) took it one step further to assert that "cognitive profiles in ASD, such as those documented here could serve as informative endophenotypes, which are key for genetic investigations" (p. 662).

However, these conclusions are problematic for many reasons. It is well documented that mean group differences in cognitive score profiles are insufficient for detecting differences among individuals (e.g., Devena and Watkins 2012; Moura et al. 2013; Styck and Watkins 2014; Ward et al. 1995; Watkins et al. 2002; Weiner 2003). As noted by

Popham (1993), “Measuring instruments that permit valid inferences about individuals will, of necessity, permit valid inferences about groups. The reverse, however, is not true” (p. 151). To mistakenly believe that differences at the group level also apply at the individual level is called the ecological fallacy (Robinson 1950).

Meehl (1989) concluded that analyses at the individual level are necessary for clinical decisions in psychology. Similar conclusions have been reached in other professions (e.g., Cotter and Peipert 2005; Metz 1978; Pfeiffer 2002). As illustrated in Fig. 1, there are four possible outcomes when test scores are used to make diagnostic decisions about individuals: true positives, false positives, true negatives, and false negatives. Thus, there are two types of errors (false positive and false negative) and two types of correct decisions (true positive and true negative). Overemphasis of true positive decisions without considering the false positive or true negative rates is a cognitive error known as pseudodiagnosticity (Doherty et al. 1979).

Several diagnostic accuracy ratios can be calculated from these four outcomes, including sensitivity, sensitivity, positive predictive power, and negative predictive power (see Fig. 1 for computational details). Unfortunately, each of these indices is influenced by the prevalence of the disorder or the cutoff score used (McFall and Treat 1999; Metz 1978). To ameliorate these limitations, receiver operating characteristic (ROC) methods have been developed (Metz 1978). The ROC provides a diagnostic accuracy index of the test that is independent of prevalence rates and cutoff scores (McFall and Treat 1999) and “is recognized widely as the most meaningful approach to quantify the accuracy of diagnostic information and diagnostic decisions” (Metz and Pan 1999, p. 1). Youngstrom (2014) provided a primer on ROC methods as they pertain to psychological assessment.



Reference Test	Diagnostic Sign +	Diagnostic Sign −
+	TP = 45	FN = 5
−	FP = 15	TN = 35

Fig. 1 Hypothetical cross-tabulation of the presence of a diagnostic sign with results of the reference test for 100 individuals (i.e., 50 cases and 50 controls). In the figure, + and – signs indicate the presence (i.e., +) and absence (i.e., −) of a diagnostic sign or the results of a reference test. Shaded boxes represent “accurate” decisions, and un-shaded boxes represent “inaccurate” decisions. *TP* true positive, *FN* false negative, *FP* false positive, *TN* true negative. Sensitivity is computed as TP / (TP + FN) = 45 / (45 + 5) = 0.90. Specificity is computed as TN / (TN + FP) = 35 / (35 + 15) = 0.70. Consequently, the false positive rate is 1 − specificity = 0.1 – 0.70 = 0.30. Positive predictive power is computed as TP / (TP + FP) = 45 / (45 + 15) = 0.75. Negative predictive power is computed as TN / (TN + FN) = 35 / (35 + 5) = 0.88

The resultant purpose of the present study was to investigate the diagnostic utility of the WISC-IV VCI and PRI > WMI and PSI cognitive score profile with ROC methods for a referred sample of school-aged students diagnosed with ASD compared to the WISC-IV standardization sample and a referred sample of children and adolescents whose school-based evaluations did not result in a diagnosis. If the VCI and PRI > WMI and PSI cognitive score profile is diagnostically useful for children and adolescents with ASD, then it should be able to distinguish between these groups. In particular, diagnostic accuracy should be highest for the comparison between children and adolescents with ASD and the WISC-IV standardization sample as the latter represents a pure sample of typically developing individuals (Wechsler 2003b). However, Meehl and Rosen (1955) cautioned that results of diagnostic utility investigations between clinical and nonclinical groups do not generalize to diagnostic decisions made for individuals “…who are referred for testing” (p. 199). Consequently, the VCI and PRI > WMI and PSI cognitive score profile must also be able to distinguish between a referred sample of children and adolescents diagnosed with ASD and an undiagnosed referred sample in order to be clinically useful for applied psychologists. Referred and clinical samples tend to share similar characteristics with lower mean scores that are more variable than the normative mean of a test (Styck and Watkins 2014; Watkins 2010). Therefore, diagnostic accuracy should be attenuated for this comparison relative to the comparison with the WISC-IV standardization sample.

Method

Participants

Clinical Group Archival special education records from approximately 7500 files in two suburban school districts located in the southwestern USA were examined for the presence of WISC-IV core battery subtest, index, and FSIQ scores upon the approval of school districts and university institutional review boards. Participants in the clinical group included 79 (78% male) children and adolescents aged 6.2 to 16.6 years ($M = 10.4$; SD = 2.7) diagnosed with ASD according to school-based evaluations. Diagnoses of ASD were made according to criteria outlined in state and US federal regulations governing special education procedures that define autism as “a developmental disability significantly affecting verbal and nonverbal communication and social interaction, generally evident before age 3, that adversely affects a child's educational performance" (Individuals with Disabilities in Education Act 2004) and further specify that "characteristics [of autism] include irregularities and impairments in communication, engagement in repetitive activities and stereotyped movements, resistance to environmental change or change in daily routines and unusual responses to sensory experiences" (Arizona Administrative Code R7-2-401 n.d.).

Approximately 43% of clinical group participants had comorbid speech-language impairment. Children and adolescents with ASD and other comorbid conditions (e.g., attention-deficit/hyperactivity disorder) were excluded from the present study to rule out the possibility of influence from other disabilities as the purpose of the present study was to determine the degree to which the cognitive score profile is unique to individuals with ASD. Approximately 82.3% of participants in the clinical sample identified as Caucasian, 8.9% identified as Hispanic, 6.3% identified as African American, and 2.5% identified as Asian/Pacific Islander.

Non-Clinical Comparison Groups Participants in the clinical group were compared to two non-clinical groups of children and adolescents without disabilities: (a) the WISC-IV standardization sample[2] and (b) a referred sample of children and adolescents whose school-based psychoeducational evaluations did not result in a diagnosis. The WISC-IV standardization sample consists of 2200 typically developing children and adolescents (50% male) aged 6 to 16.9 years ($M = 11.5$; SD = 3.2) specifically excluding (a) individuals with uncorrected visual and hearing impairment; (b) those not fluent in English; (c) primarily non-verbal individuals; (d) those with upper extremity disabilities; (e) current admittance to hospital, mental, or psychiatric facility; (f) children currently taking medication that might affect performance; and (g) any child who had a previous diagnosis of any physical condition or illness (i.e., stroke, epilepsy, brain tumor; Wechsler 2003b). Participants in the undiagnosed referred sample included 216 children and adolescents (64.8% male) aged 6.1 to 16.7 years ($M = 9.7$; SD = 2.2) without educational disabilities according to results of school-based evaluations. Approximately 80.6% of participants in the undiagnosed referred sample identified as Caucasian, 7.4% identified as Hispanic, 4.6% identified as African American, 4.6% identified as American Indian, and 1.4% identified as Asian/Pacific Islander.

Procedure

The fourth technical report issued by the WISC-IV publishers introduced a sixth composite score purportedly "sensitive to cases in which working memory performance is discrepant from verbal comprehension performance and/or processing speed performance is discrepant from perceptual reasoning performance" referred to as the general ability index (GAI; Raiford et al. 2005, p. 2), which is comprised of the core subtests that form the VCI and PRI. Likewise, Weiss and Gabel (2008) advocated use of the cognitive proficiency index (CPI) for summarizing performance on the subtests that form the WMI and PSI. Consequently, the VCI and PRI > WMI and PSI cognitive score profile is equivalent to a cognitive score profile in which the GAI > CPI. To determine the presence of the GAI > CPI profile, participants' PRI and VCI scores were transformed into the GAI and participants' WMI and PSI scores were transformed into the CPI using the tables in Weiss et al. (2008).

Participants' diagnostic status as per school-based evaluations served as the diagnostic reference and the degree to which participants' WISC-IV GAI scores were greater than participants' CPI scores served as the index test. First, diagnostic status was cross-tabulated with the presence/absence of the GAI > CPI cognitive score profile to produce frequencies of true positive, false positive, true negative, and false negative decisions. Next, a difference score was computed by subtracting the CPI from the GAI. If the GAI > CPI cognitive score profile is able to accurately distinguish participants with and without ASD, then larger GAI-CPI difference scores should be associated with an increased probability of belonging to the ASD participant subgroup.

Instrument

The WISC-IV is an individually administered intelligence test that was normed on 2200 children and adolescents 6–16 years old who were stratified into groups based on age, sex, race, ethnicity, geographic region, and parent education level to match the 2000 US census. It is comprised of 10 core subtests ($M = 10$; SD = 3) that combine to form the FSIQ and four index scores ($M = 100$; SD = 15): the VCI, the PRI, the WMI, and the PSI. The Similarities, Vocabulary, and Comprehension subtests form the VCI, which is a measure of verbal reasoning. The PRI measures non-verbal reasoning and visual-spatial skills. It is formed from the Block Design, Picture Concepts, and Matrix Reasoning subtests. The WMI is formed from the Digit Span and Letter-Number Sequencing subtests and is a measure of working memory. The PSI is formed from the Coding and Symbol Search subtests and is a measure of processing speed. Reliability and validity of WISC-IV test scores fall within acceptable limits for applied use (Flanagan and Kaufman 2004; Wechsler 2003b). For example, Wechsler (2003b) reported reliability coefficients of 0.94, 0.92, 0.92, and 0.88 for the VCI, PRI, WMI, and PSI scores, respectively.

---

[2] Standardization data from the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV). Copyright © 2003 NCS Pearson, Inc. Used with permission. All rights reserved.

Analyses

All analyses were conducted with *R* version 3.2.0 (R Core Development Team 2013). First, one-way analysis of variance (ANOVA) was used to evaluate the presence of mean differences between participant subgroups for the WISC-IV core battery subtests, index scores, and FSIQ score. The Welch (1947) approximate *F* test was referenced to evaluate the omnibus ANOVA to relax the assumption of equal variances. The Bonferonni correction procedure was applied to evaluate all pairwise differences to control for inflation of type I error and maintain an experimentwise alpha rate of 0.05 (i.e., 0.05/51 = 0.001). Approximately 52 participants per group are required to detect a medium group effect, and 14 participants per group are required to detect a strong group effect to maintain a minimum statistical power of 0.80 (Cohen 1988).

Next, coordinates of true positive (i.e., sensitivity, *y*-axis) and false positive (i.e., 1 − specificity, *x*-axis) decisions derived from cross-tabulating diagnostic status with the presence/absence of the GAI > CPI cognitive score profile were plotted on a scatterplot for all possible decision thresholds (i.e., the value obtained from subtracting the CPI from the GAI; Pepe 2003). Connecting the *x*- and *y*-coordinates of true positive and false positive rates on the scatterplot produce a ROC curve. The area under the curve (AUC) provides an index of the overall accuracy of a diagnostic sign (Fawcett 2006; Pepe 2003), and the accuracy of a diagnostic sign improves from chance probability (i.e., 50%, flipping a fair coin) when true positive rates increase (*y*-axis) and false positive rates decrease (*x*-axis). In other words, the accuracy of a diagnostic sign improves as the AUC increases. In the context of the present study, the AUC is interpreted as the probability that a randomly selected individual from the clinical participant group displays a larger difference between their GAI and CPI scores than a randomly selected individual from the non-clinical comparison group.

The AUC was computed using a non-parametric statistical approach (Pintea and Moldovan 2009). Non-parametric and parametric methods for computing the AUC tend to yield similar results, but the non-parametric approach does not rely on strict distributional assumptions (Hajian-Tilaki et al. 1997) that are often unmet in referred or clinical samples (Styck and Watkins 2014; Watkins 2010). AUC values range from 0.00 to 1.00, with a value of 0.50 representing chance probability. AUC values between 0.50 and 0.70 have been suggested to characterize low accuracy; AUC values between 0.70 and 0.90 have been suggested to characterize moderate accuracy, and AUC values between 0.90 and 1.00 have been suggested to characterize high accuracy (Streiner and Cairney 2007; Swets 1996). Alternatively, a "good" AUC value might be established in comparison to AUC values obtained from alternative testing methods (Kraemer 1992). Using that standard, Youngstrom (2014) reported that AUC estimates ranging from 0.70 to 0.80 have been demonstrated for behavioral checklists and inventories. According to the formulas provided by Obuchowski et al. (2004), approximately 7–31 participants per group are required to detect an AUC value within the medium accuracy range and approximately 3–7 participants per group are required to detect an AUC value within the high accuracy range in order to maintain a minimum statistical power of 0.80.

Results

Results indicated statistically significant differences between participant subgroups on the VCI, $F(2, 168.8) = 13.1$; $p < 0.001$, the WMI, $F(2, 168.3) = 33.0$; $p < 0.001$, the PSI, $F(2, 168.4) = 35.6$; $p < 0.001$, the CPI, $F(2, 168.6) = 43.4$; $p < 0.001$, the FSIQ, $F(2, 169.9) = 22.7$; $p < 0.001$, and the GAI-CPI difference score, $F(2, 165.5) = 17.6$; $p < 0.001$. Post hoc pairwise comparisons indicated that mean FSIQ scores for both the ASD ($d = 0.53$; $p < 0.01$) and undiagnosed referred samples ($d = 0.34$; $p < 0.001$) were significantly lower than the WISC-IV standardization sample. In addition, mean scores for the ASD and undiagnosed referred samples were significantly lower than the WISC-IV standardization sample on the CPI ($d = 0.79$; $p < 0.001$ and $d = 0.47$; $p < 0.001$, respectively), PSI ($d = 0.93$; $p < 0.001$ and $d = 0.28$; $p < 0.01$, respectively), and WMI ($d = 0.45$; $p < 0.05$ and $d = 0.49$; $p < 0.001$, respectively). Mean PSI scores for the ASD sample were also significantly lower than mean PSI scores for the undiagnosed referred sample ($d = 0.75$; $p < 0.001$). In addition, the undiagnosed referred sample exhibited significantly lower mean scores on the VCI ($d = 0.29$; $p < 0.001$). Finally, the mean GAI-CPI difference score for both the ASD and undiagnosed referred samples were significantly higher than the WISC-IV standardization sample ($d = 0.55$; $p < 0.01$ and $d = 0.30$; $p < 0.01$, respectively).

At the subtest level, statistically significant differences between participant subgroups emerged on the Digit Span, $F(2, 168.0) = 32.4$; $p < 0.001$, Coding, $F(2, 166.6) = 32.4$; $p < 0.001$, Vocabulary, $F(2, 167.2) = 14.5$; $p < 0.001$, Letter-Number Sequencing, $F(2, 166.8) = 17.1$; $p < 0.001$, Comprehension, $F(2, 168.5) = 17.1$; $p < 0.001$, and Symbol Search subtests, $F(2, 168.6) = 22.9$; $p < 0.001$. Post hoc pairwise comparisons indicated that mean scores on the Coding subtest were significantly lower for the ASD group ($d = 0.88$; $p < 0.001$) and undiagnosed referred group ($d = 0.36$; $p < 0.001$) when compared to the WISC-IV standardization sample. In addition, mean Coding subtest scores were lower for the ASD group when compared to the undiagnosed referred group ($d = 0.61$; $p < 0.01$). Participants in the undiagnosed referred group exhibited significantly lower mean scores than the WISC-IV standardization sample on the Digit Span ($d = 0.50$; $p < 0.001$) and Vocabulary ($d = 0.36$;

$p < 0.001$) subtests. Mean scores on the Comprehension and Letter-Number Sequencing subtests were significantly lower for both the ASD group ($d = 0.57$; $p < 0.01$ and $d = 0.45$; $p < 0.05$, respectively) and the undiagnosed referred group ($d = 0.23$; $p < 0.01$ and $d = 0.28$; $p < 0.001$, respectively) when compared to the WISC-IV standardization sample. Moreover, mean scores on the Comprehension subtest were also significantly lower for the ASD group when compared to the undiagnosed referred group ($d = 0.44$; $p < 0.05$). Finally, the ASD group exhibited significantly lower mean scores on the Symbol Search subtest than either the WISC-IV standardization sample ($d = 0.77$; $p < 0.001$) or the undiagnosed referred group ($d = 0.14$; $p < 0.001$). Table 1 contains WISC-IV subtest, index, and FSIQ score means and standard deviations for all participants disaggregated by subgroup. Significant differences between groups are indicated at the $p < 0.05$, $p < 0.01$, and $p < 0.001$ levels.

Next, true positive and false positive rates were computed to determine the degree to which individuals in the ASD sample uniquely exhibited the GAI > CPI cognitive score profile. Approximately 73.4% of the ASD sample displayed a GAI > CPI. However, so did 50.2% of the WISC-IV standardization sample. Results of ROC curve analyses (see Fig. 2) yielded an AUC of 0.660, 95% CI (0.591, 0.729) indicating that a randomly selected individual from the ASD sample would exhibit a larger difference between their GAI and CPI scores than a randomly selected individual from the WISC-IV standardization sample 66.0% of the time. This value falls within the low-moderate accuracy range according to conventional benchmarks (Streiner and Cairney 2007; Swets 1996) and below the range typically found with behavioral checklists and inventories (Youngstrom 2014) as depicted in Fig. 2. Figure 3 illustrates the considerable overlap in GAI-CPI difference scores between the two participant groups. As depicted in the figure, GAI-CPI difference scores emerge in the ASD and WISC-IV standardization sample at similar rates.

There were significantly more males than females in the ASD sample when compared to the WISC-IV standardization sample, $\chi^2(1) = 23.6$; $p < 0.001$. Furthermore, differences in mean subtest and index scores have been observed for

Table 1 Means and standard deviations of WISC-IV subtest, index, and FSIQ scores for participants disaggregated by study subgroup

WISC-IV score	ASD ($n = 79$)		Undiagnosed referred ($n = 216$)	
	M	SD	*M*	SD
BD	9.9	3.7	9.8	2.9
SI	10.0	3.0	9.4	2.7
DS	8.8	3.5	8.6^{***}	2.6
PCn	9.0	3.4	10.0	2.5
CD	7.0^{***}_{**}	3.8	9.0^{***}	2.6
VC	9.1	3.8	9.0^{***}	2.5
LN	8.5^{*}	3.7	9.2^{***}	2.6
MR	9.8	3.5	9.8	2.6
CO	7.9^{**}	4.3	9.4^{**}	2.2
SS	7.6^{***}_{***}	3.2	9.6	2.6
VCI	94.4	19.7	95.5^{***}	11.4
PRI	97.3	18.9	99.7	12.5
WMI	92.1^{*}	18.2	93.0^{***}	11.8
PSI	84.9^{***}_{***}	17.5	96.3^{**}	12.4
GAI	95.5	19.0	97.7	11.2
CPI	86.9^{***}	17.8	93.6^{***}	11.7
GAI – CPI	8.61^{**}	16.6	4.1^{**}	11.3
FSIQ	91.0^{**}	18.3	95.4^{***}	11.5

Superscripts indicate significant differences between the ASD and undiagnosed referred participant subgroups and the WISC-IV standardization sample, whereas subscripts indicate significant differences between the ASD participant subgroup when compared to the undiagnosed referred participant subgroup

WISC-IV Wechsler Intelligence Scale for Children—Fourth Edition, *ASD* autism spectrum disorder, *BD* block design, *SI* similarities, *DS* digit span, *PCn* picture concepts, *CD* coding, *VC* vocabulary, *LN* letter-number sequencing, *MR* matrix reasoning, *CO* comprehension, *SS* symbol search, *VCI* verbal comprehension index, *PRI* perceptual reasoning index, *WMI* working memory index, *PSI* processing speed index, *GAI* general ability index, *CPI* cognitive processing index, *FSIQ* full-scale intelligence quotient

*$p < 0.05$; *$p < 0.01$; *$p < 0.001$

individuals with FSIQ scores > 70 compared to those with FSIQ scores ≤ 70 as previously noted (Foley-Nicpon et al. 2012; Mayes and Calhoun 2008). Consequently, additional analyses were conducted to determine the degree to which the GAI > CPI cognitive score profile could accurately distinguish between individuals in the ASD sample and individuals in the WISC-IV standardization sample matched on gender and FSIQ score. The diagnostic utility of the GAI > CPI profile did not improve. Results yielded an AUC value of 0.619, 95% CI (0.531, 0.707) indicating that a randomly selected individual from the ASD sample would exhibit a larger difference between their GAI and CPI scores than a randomly selected individual from the WISC-IV standardization sample 61.9% of the time after controlling for differences in gender and overall cognitive ability (see Fig. 2). An AUC value of this magnitude has been characterized as having low-moderate accuracy according to conventional benchmarks (Streiner and Cairney 2007; Swets 1996) and below the range typically found with behavioral checklists and inventories (Youngstrom 2014) as depicted in Fig. 2.

Approximately 61.1% of the undiagnosed referred sample displayed a GAI > CPI cognitive score profile, denoting even greater overlap between these two participant subgroups. Results of ROC curve analyses (see Fig. 2) yielded an AUC of 0.598, 95% CI (0.519, 0.678), which is low accuracy according to conventional benchmarks (Streiner and Cairney 2007; Swets 1996) and is below the range typically found with behavioral checklists and inventories (Youngstrom 2014). This indicates that a randomly selected individual from the ASD sample would exhibit a larger difference between their GAI and CPI scores than a randomly selected individual from the undiagnosed referred sample 59.8% of the time. However, power to detect an AUC value at this level requires a greater number of participants per group than were included in the present study. Power for this comparison dropped slightly below 0.80 ($1-\beta = 0.76$).

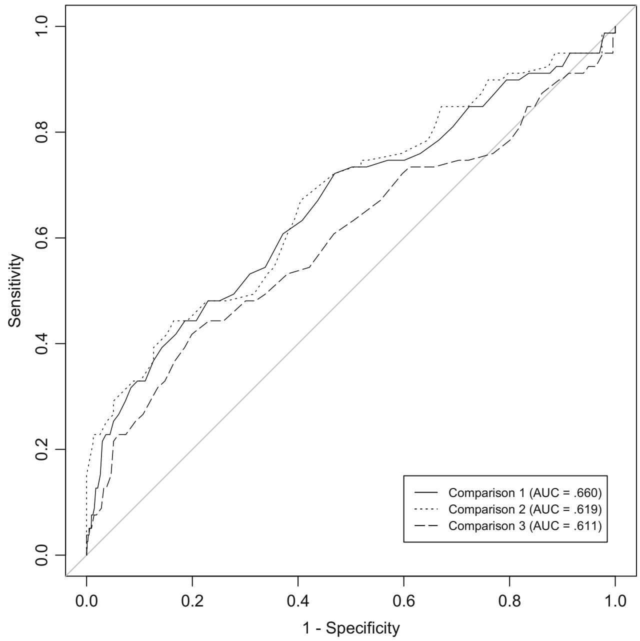


Fig. 2 Graph depicting true positive (i.e., sensitivity) and false positive (i.e., 1 − specificity) rates for three ROC curves. Comparison 1 consisted of a sample of children and adolescents with ASD ($n = 79$) compared to the WISC-IV standardization sample ($n = 2200$). Comparison 2 consisted of a sample of children and adolescents with ASD ($n = 79$) compared to WISC-IV standardization sample participants matched on gender and FSIQ score ($n = 79$). Comparison 3 consisted of a sample of children and adolescents with ASD ($n = 79$) compared to a referred sample of children and adolescents whose school-based evaluations did not result in a diagnosis ($n = 216$)

Discussion

The purpose of the present study was to determine the degree to which the WISC-IV GAI > CPI cognitive score profile could accurately distinguish between a sample of children and adolescents with ASD and two non-clinical comparison groups, the WISC-IV standardization sample and a referred sample of children and adolescents whose school-based evaluations did not result in a diagnosis. Overall, mean differences that emerged in the present study generally replicated those reported previously on the cognitive score profiles of the WISC-IV for children and adolescents with ASD. The mean PSI score for the ASD group was significantly lower than the WISC-IV standardization sample. This finding has been consistently documented in the extant literature, and effect sizes have been similar across studies ($d = 0.78$–0.93; Mayes and Calhoun 2008; Oliveras-Rentas et al. 2012). The mean WMI score for the ASD group in the present study was not significantly lower than the WISC-IV standardization sample as was reported in Mayes and Calhoun (2008). However, the mean CPI score (i.e., formed from the core subtests that comprise the PSI and WMI) for the ASD group was significantly lower than the WISC-IV standardization sample, and no significant mean strengths on the VCI or PRI emerged when comparing the ASD group with the WISC-IV standardization sample as was previously reported in Mayes and Calhoun (2008).

At the subtest level, mean scores on Coding and Symbol Search were significantly lower for the ASD group than the WISC-IV standardization sample. This pattern has been documented in prior studies with similar effect sizes noted ($d = 0.80$–0.96 for Coding and $d = 0.59$–0.77 for Symbol Search; Mayes and Calhoun 2008; Oliveras-Rentas et al. 2012). Mean scores for the ASD group were also significantly lower on the Comprehension subtest when compared to the WISC-IV standardization sample as was reported in Oliveras-Rentas et al. (2012). Effect sizes for this pattern were comparable across studies ($d = 0.57$–0.61).

The GAI > CPI cognitive score profile must be unique to individuals with ASD in order for it to be diagnostically useful for applied psychologists. Moreover, psychological evaluations are not administered to the general population—they are administered to individuals who are referred for testing. Consequently, mean differences in WISC-IV subtest, index, and FSIQ scores were also evaluated between the ASD group and a referred sample of children and adolescents whose school-based evaluations did not result in a diagnosis. Mean scores for the ASD group were statistically lower than mean scores for the undiagnosed referred group on the PSI and the two subtests that form the PSI, Symbol Search, and Coding. However, mean scores on the WMI for the undiagnosed referred sample were significantly lower than the WISC-IV standardization sample along with mean scores on the Digit Span subtest, and this pattern was not observed when the ASD group was compared to the WISC-IV standardization sample. Furthermore, mean scores on the CPI (i.e., formed from the core subtests that comprise the PSI and WMI) were significantly lower for both the ASD group and the undiagnosed referred group when compared to the WISC-IV standardization sample. These group trends suggest that most children and adolescents who are referred for a psychological evaluation exhibit deficits in working memory and speed of information processing as measured by a commonly administered standardized intelligence test, regardless of whether or not those evaluations eventuate in a diagnosis. As a result, this pattern of performance is not unique to individuals with ASD as a group.

Most importantly, statistically significant group trends were insufficient for making accurate predictions about individual performance. Results of diagnostic utility statistics indicated that the GAI > CPI cognitive score profile distinguished between individuals in the ASD group and the WISC-IV standardization sample or the undiagnosed referred sample at slightly better than chance rates (AUC = 0.598–0.660) across all possible thresholds despite the observation of statistically significant mean differences between groups on the CPI with a medium-large effect sizes (*d* = 0.47–0.79). This is due to the large amount of distributional overlap between groups. Figure 3 illustrates this overlap for the ASD group and the WISC-IV standardization sample with respect to the difference between participants' GAI and CPI scores. Although 73.4% of individuals in the ASD group exhibited the GAI > CPI cognitive score profile, so did 50.2–61.1% of individuals in the non-clinical comparison groups. This is an unacceptably high false positive rate for individual decisions. In addition, positive predictive power indicates that the probability of having ASD, given a GAI > CPI profile, was only 5% when individuals in the ASD group were compared to the WISC-

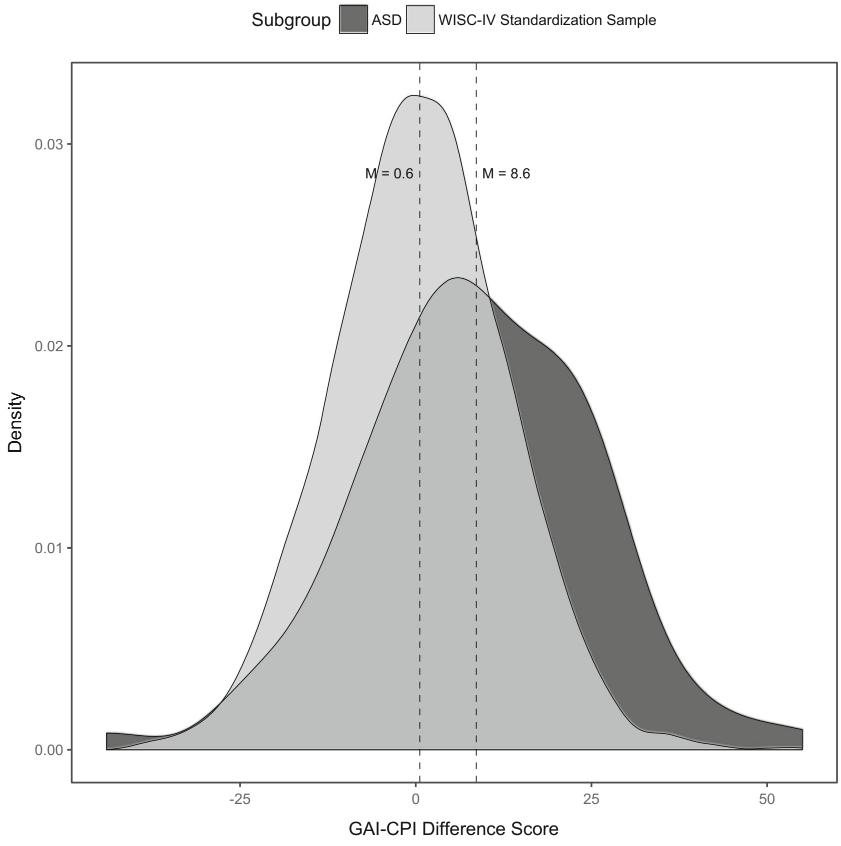


Fig. 3 Histogram depicting the distributional overlap of GAI-CPI difference scores for a sample of 79 children and adolescents with school-based diagnoses of ASD and the 2200 children and adolescents contained in the WISC-IV standardization sample. Diagonal lines represent the mean GAI-CPI difference score for each group. Mean GAI-CPI difference scores were significantly different between the ASD and WISC-IV standardization samples with a medium effect size ($d = 0.55$; $p < 0.01$)

IV standardization sample. This value improved to 55.8% (slightly above chance rates) after controlling for differences in gender and FSIQ score. However, positive predictive power declined to 30.5% when individuals in the ASD group were compared to individuals in the undiagnosed referred group.

Results of the present study echo those of countless others that have documented the poor diagnostic utility of cognitive score profiles (e.g., Devena and Watkins 2012; Moura et al. 2013; Styck and Watkins 2013, 2014; Ward et al. 1995; Watkins et al. 2002). “Mean differences between two groups obtained in nomothetic research, even when statistically significant by usual standards, rarely have sufficient predictive power to support reliable inferences in idiographic appraisals, that is, in deciding whether a particular person should be classified as belonging to one group or another. Psychologists who uncritically use a group difference as a diagnostic criterion are losing sight of basic statistical realities” (Weiner 2003, p. 336). Unfortunately, the notion that cognitive score profiles identified from group trends can be used to inform decisions for individuals continues to be perpetuated in easily accessible formats for psychologists who work in applied settings, such as in books, book chapters, and technical manuals that accompany individually administered intelligence test batteries (e.g., Flanagan et al. 2013; Foley-Nicpon et al. 2012; Mayes and Calhoun 2008; Oliveras-Rentas et al. 2012; Wechsler 2003b; Weiss et al. 2016) despite mounting empirical evidence refuting these practices published in scholarly journals.

The present study is not without limitations. Diagnoses of ASD resulted from school-based evaluations, and participants may differ from children and adolescents diagnosed with ASD using other criteria. However, results may generalize to school-based samples, and participants in the ASD group resembled those of individuals diagnosed with ASD nationwide on demographic characteristics such as gender and ethnicity (CDC 2014). A second limitation is the use of the WISC-IV, which has been replaced by the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V; Wechsler 2014). ASD is a low incidence disability. The majority of research on the cognitive score profiles of individuals with ASD has been conducted on outdated tests due to the difficulty of recruiting large samples (Cederlund and Gillberg 2004; Klin et al. 2005; Planche and Lemonnier 2012; Zander and Dahlgren 2010). Moreover, results of the present study are consequential in that they demonstrate the negative impact of overreliance on group trends in deriving frameworks for guiding diagnostic decisions and suggest a need to reanalyze the literature conducted on the cognitive score profiles of individuals with ASD at the individual level. The present study marks the first to do so for individuals with ASD to date. Finally, observed scores contain measurement error, and the standard error of measurement (SEM) is used to compute confidence intervals that create upper and lower bounds within which true scores are likely located. To date, the WISC-IV test publisher has not reported the SEM for the GAI or CPI composite scores or the SEM for GAI-CPI difference score. This information could assist independent researchers in evaluating the clinical significance of score discrepancies at the group and individual levels. Future iterations of the Wechsler scales should report the SEM for all scores that are derived from the test and used to draw inferences about a person’s cognitive ability.

While it is true that psychological assessment rarely, if ever, solely involves reference to a single diagnostic sign, to paraphrase Meehl and Rosen (1955), all psychological evaluations eventuate in a diagnostic decision. The accuracy of those decisions depends upon the diagnostic utility of the information we gather, whether it comes from standardized individually administered intelligence tests or clinical interviews. Results of the present study indicate that group trends in the cognitive score profiles of individuals with ASD suggesting relative strengths in verbal and non-verbal reasoning and relative weaknesses in working memory and speed of information processing are inaccurate predictors of individual performance. In the absence of idiographic information about cognitive profiles, psychologists who work in applied settings “are well-advised to proceed cautiously, if at all, in basing predictive or postdictive conclusions on it in the individual case” (Weiner 2003, p. 336).

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Arizona Administrative Code 7 A.A.C. § 2-4 n.d..

Barnhill, G., Hagiwara, T., Myles, B. S., & Simpson, R. L. (2000). Asperger syndrome: A study of the cognitive profiles of 37 children and adolescents. *Focus on Autism and Other Developmental Disabilities, 15*, 146–153.

Braden, J. P. (2013). Psychological assessment in school settings. In J. A. Naglieri & J. R. Graham (Eds.), *Handbook of psychology: Assessment psychology* (Vol. 10, 2nd ed., pp. 261–290). Hoboken: John Wiley & Sons, Inc..

Calhoun, S. L., & Mayes, S. D. (2005). Processing speed in children with clinical disorders. *Psychology in the Schools, 42*, 333–343. https://doi.org/10.1002/pits.20067

Cederlund, M., & Gillberg, C. (2004). One hundred males with Asperger syndrome: a clinical study of background and associated factors. *Developmental Medicine and Child Neurology, 46*, 652–660.

Centers for Disease Control and Prevention. (2014). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2010 (Surveillance Summaries No. 63–2). Atlanta: U.S. Department of Health and Human Services.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences—second edition. Mahwah: Lawrence Erlbaum & Associates.

Cotter, K., & Peipert, J. F. (2005). Can you handle the truth (and know it when you see it)? Understanding sensitivity, specificity, predictive values, and ROC curves. *Journal of Minimally Invasive Gynecology, 12*, 385–390. https://doi.org/10.1016/j.jmig.2005.05.021

Devena, S. E., & Watkins, M. W. (2012). Diagnostic utility of WISC-IV general abilities index and cognitive proficiency index difference scores among children with ADHD. *Journal of Applied School Psychology, 28*, 133–154. https://doi.org/10.1080/15377903.2012.669743

Doherty, M. E., Mynatt, C. R., Tweney, R. D., & Schiavo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica, 43*, 111–121.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 2*, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken: Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). Essentials of cross-battery assessment—third edition. Hoboken: John Wiley & Sons.

Foley-Nicpon, M., Assouline, S. G., & Stinson, R. D. (2012). Cognitive and academic distinctions between gifted students with autism and asperger syndrome. *Gifted Child Quarterly, 56*, 77–89. https://doi.org/10.1177/0016986211433199

Hajian-Tilaki, K. O., Hanley, J. A., Joseph, L., & Collet, J. P. (1997). A comparison of parametric and nonparametric approaches to ROC analysis of quantitative diagnostic tests. *Decision Making, 17*, 94–102.

Happe, F. G. E. (1994). Wechsler IQ profile and theory of mind in autism: a research note. *Journal of Child Psychology and Psychiatry, 35*, 1461–1471.

Individuals with Disabilities Education Improvement Act. (2004). Public Law 108–446, 108th Cong., 118 Stat. 2647 (enacted).

Klin, A., Pauls, D., Shultz, R., & Volkmar, F. (2005). Three diagnostic approaches to Asperger syndrome: implications for research. *Journal of Autism and Developmental Disorders, 35*, 221–234. https://doi.org/10.1007/s10803-004-2001-y

Kraemer, H. C. (1992). Evaluating medical tests–objective and quantitative guidelines. Newbury Park: Sage.

Koyama, T., Tachimori, H., Osada, H., Takeda, T., & Kurita, H. (2007). Cognitive and symptom profiles in Asperger's syndrome and high-functioning autism. *Psychiatry and Clinical Neurosciences, 61*, 99–104. https://doi.org/10.1111/j.1440-1819.2007.01617.x

Lincoln, A. J., Courchesne, E., Kilman, B. A., Elmasian, R., & Allen, M. (1988). A study of intellectual abilities in high-functioning people with autism. *Journal of Autism and Developmental Disorders, 18*, 505–524.

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

Mayes, S. D., & Calhoun, S. L. (2008). WISC-IV and WIAT-II profiles in children with high-functioning autism. *Journal of Autism and Developmental Disorders, 38*, 428–439. https://doi.org/10.1007/s10803-007-0410-4

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50*, 215–241. https://doi.org/10.1146/annurev.psych.50.1.215

Meehl, P. E. (1989). Law and the fireside inductions (with postscript): some reflections of a clinical psychologist. *Behavioral Sciences & the Law, 7*, 521–550. https://doi.org/10.1002/bsl.2370070408

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, and cutting scores. *Psychological Bulletin, 52*, 194–216. https://doi.org/10.1037/h0048070

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8*, 283–298. https://doi.org/10.1016/S0001-2998(78)80014-2

Metz, C. E., & Pan, X. (1999). "Proper" binormal ROC curves: theory and maximum-likelihood estimation. *Journal of Mathematical Psychology, 43*, 1–33. https://doi.org/10.1006/jmps.1998.1218

Moura, O., Simões, M. R., & Pereira, M. (2013). WISC-III cognitive profiles in children with developmental dyslexia: specific cognitive disability and diagnostic utility. *Dyslexia, 20*, 19–37. https://doi.org/10.1002/dys.1468

Obuchowski, N. A., Lieber, M. L., & Wians, F. H. (2004). ROC curves in clinical chemistry: uses, misuses, and possible solutions. *Clinical Chemistry, 50*, 1118–1125. https://doi.org/10.1373/clinchem.2004.031823

Oliveras-Rentas, R. E., Kenworthy, L., Roberson, R. B., Martin, A., & Wallace, G. L. (2012). WISC-IV profile in high-functioning autism spectrum disorders: impaired processing speed is associated with increased autism communication symptoms and decreased adaptive communication abilities. *Journal of Autism and Developmental Disorders, 42*, 655–664. https://doi.org/10.1007/s10803-011-1289-7

Pepe, M. S. (2003). *Statistical evaluation of medical tests for classification and prediction*. New York: Oxford University Press.

Pfeiffer, D. U. (2002). *Veterinary epidemiology: an introduction*. London: Royal Veterinary College, University of London.

Pintea, S., & Moldovan, R. (2009). The receiver-operating characteristics (ROC) analysis: fundamentals and applications of clinical psychology. *Journal of Cognitive and Behavioral Psychotherapies, 9*, 49–66.

Planche, P., & Lemonnier, E. (2012). Children with high-functioning autism and asperger's syndrome: can we differentiate their cognitive profiles? *Research in Autism Spectrum Disorders, 6*, 939–948. https://doi.org/10.1016/j.rasd.2011.12.009

Popham, W. J. (1993). *Educational evaluation* (3rd ed.). Boston: Allyn and Bacon.

R Core Team. (2013). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing Retrieved from: http://www.R-project.org/

Raiford, S. E., Weiss, L. G., Rolfhus, E. L., & Coalson, D. (2005). *Wechsler intelligence scale for children-fourth edition, general ability index (Technical Report No. 4)*. San Antonio: Harcourt Assessment.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15*, 351–357. https://doi.org/10.2307/2087176

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristic curves. *Canadian Journal of Psychiatry, 52*, 121–128.

Styck, K. M., & Watkins, M. W. (2013). Diagnostic utility of the culture-language interpretive matrix for the Wechsler Intelligence Scales for Children—Fourth Edition among referred students. *School Psychology Review, 42*, 367–382.

Styck, K. M., & Watkins, M. W. (2014). Discriminant validity of the WISC-IV culture-language interpretive matrix. *Contemporary School Psychology, 18*, 168–177. https://doi.org/10.1007/s40688-014-0021-y

Swets, J. A. (1996). *Signal detection theory and roc analysis in psychology and diagnosis: collected papers*. Mahwah: Erlbaum.

Ward, S. B., Ward, T. J., Hatt, C. V., Young, D. L., & Mollner, N. R. (1995). The incidence and utility of the ACID, ACIDS, and SCAD profiles in a referred population. *Psychology in the Schools, 32*, 267–276. https://doi.org/10.1002/1520-6807

Watkins, M. W. (2010). Structure of the Wechsler Intelligence Scale for Children—Fourth Edition among a national sample of referred students. *Psychological Assessment, 22*, 782–787. https://doi.org/10.1037/a0020043

Watkins, M. W., Kush, J. C., & Schaefer, B. A. (2002). Diagnostic utility of the learning disability index. *Journal of Learning Disabilities, 35*, 98–103.

Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children—Fourth edition*. San Antonio: The Psychological Corporation.

Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children—Fourth Edition technical manual*. San Antonio: The Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children—Fifth Edition*. San Antonio: The Psychological Corporation.

Weiner, I. B. (2003). Prediction and postdiction in clinical decision making. *Clinical Psychology: Science and Practice, 10*, 335–338. https://doi.org/10.1093/clipsy/bpg030

Weiss, L. G., & Gabel, A. D. (2008). *Using the cognitive proficiency index in psychoeducational assessment (Technical Report No. 6)*. San Antonio: Harcourt Assessment.

Weiss, L. G., Beal, A. L., Saklofske, D. H., Alloway, T. P., & Prifitera, A. (2008). Interpretation and intervention with WISC-IV in the clinical assessment context. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical assessment and intervention—second edition* (pp. 3–66). San Diego: Elsevier Inc..

Weiss, L. G., Saklofske, D. H., Holdnack, J., & Prifitera, A. (2016). *WISC-V clinical use and interpretation*. San Diego: Elsevier/ Academic Press.

Welch, B. L. (1947). The generalization of students' problem when several different population variances are involved. *Biometrika, 34*, 28–35.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *Journal of Pediatric Psychology, 39*, 204–221. https://doi.org/10.1093/jpepsy/jst062

Zander, E., & Dahlgren, S. V. (2010). WISC-III index score profiles of 520 swedish children with pervasive developmental disorders. *Psychological Assessment, 22*, 213–222. https://doi.org/10.1037/a0018335

Kara M. Styck received her Ph.D. in school psychology from Arizona State University and she is currently an assistant professor in the Educational Psychology Department at the University of Texas at San Antonio. Her research interests include individual differences in cognitive and affective factors that impact success in school with a particular emphasis in the psychometrics of assessment and decision-making.

Miriam S. Aman received her M.A. in school psychology from the University of Texas at San Antonio. She is currently a Licensed Specialist in School Psychology at Dallas Independent School District. She primarily works with students with autism spectrum disorder and emotional disturbance. Miriam enjoys working with teachers to implement behavior management strategies in the classroom.

Marley W. Watkins received his Ph.D. in school psychology from the University of Nebraska-Lincoln and hasheld positions with the Deer Valley Unified School District, Pennsylvania State University, Arizona State University, and Baylor University. He is currently Non-Resident Scholar in the Department of Educational Psychology at Baylor University. His research interests include professional issues, the psychometrics of assessment and diagnosis, individual differences, and computer applications. Dr. Watkins has published more than 150 peer-reviewed journal articles and made more than 100 presentations at professional conferences.