J Autism Dev Disord (2016) 46:3183–3189
DOI 10.1007/s10803-016-2851-0

BRIEF REPORT

# Brief Report: Parent-Teacher Discrepancies on the Developmental Social Disorders Scale (BASC-2) in the Assessment of High-Functioning Children with ASD

**Christopher Lopata[1] · James P. Donnelly[1] · Allyson K. Jordan[2] · Marcus L. Thomeer[1] · Christin A. McDonald[1] · Jonathan D. Rodgers[1]**

Published online: 11 July 2016


**Abstract** This study compared parent and teacher ratings of ASD-related symptoms of 120 high-functioning children, ages 6–12 years with ASD (HFASD) using the Developmental Social Disorders (DSD) scale of the BASC-2. DSD ratings (parent and teacher) were significantly higher than normative estimates. The cross-informant comparison was significantly higher for parents (vs. teachers), and correlations (ICC and Pearson) between the informant groups were significant (but low in magnitude). Agreement among parents and teachers accurately placed 81 % of cases above the at-risk cutpoint for symptoms of ASD, and agreement was highest in the at-risk range of perceived symptoms. Additional analyses indicated a significant difference in the trend across the parent-teacher discrepancies, and no significant moderators of the discrepancies. Implications for assessment are provided.

**Keywords** Informant discrepancies · ASD-related symptoms · BASC-2 DSD · Screening

## Introduction

High-functioning children with autism spectrum disorder (HFASD) exhibit core symptoms involving social-communication impairments and circumscribed and repetitive behaviors and interests, but are differentiated based on relatively intact cognitive and language abilities (APA 2013). Children with HFASD comprise nearly half of children with ASD and data suggest that this subgroup may account for the recent increase in ASD prevalence (CDC 2014). The increase in children with HFASD indicates a growing need for measures that help with identification and outcome assessments.

Rating scales are commonly administered to identify children in need of a comprehensive ASD assessment and/or inform diagnostic decisions. They are inexpensive, easy to administer, brief, and based on ratings from informants in naturalistic settings, and allow for ratings from multiple informants (Constantino and Gruber 2012; Norris and Lecavalier 2010). Many utilize continuous scaling which is consistent with the diagnostic framework that conceptualizes ASD symptoms and impairments on a continuum of severity (APA 2013). Rating scales also allow for graded measurement and where the child falls within a clinical category (Achenbach 2011). Lastly, measuring symptom severity on a continuous scale makes rating scales useful in monitoring treatment response (Reynolds and Kamphaus 2004).

Clinical assessments require consideration and integration of information from multiple informants. Parents and teachers are critical sources (Norris and Lecavalier 2010), yet studies have generally found low-to-moderate correlations between parents' and teachers' ratings of typically-developing youth (e.g., Achenbach et al. 1987). These findings have led to recommendations for empirical studies of informant discrepancies (De Los Reyes 2011) as they can affect diagnostic decisions, treatment access, and progress monitoring (De Los Reyes and Kazdin 2004). To date, little research has comprehensively examined informant agreement and reliability for ratings of ASD-related

✉ Christopher Lopata
lopatac@canisius.edu

[1] Canisius College, Institute for Autism Research, Science Hall 1016C, 2001 Main Street, Buffalo, NY 14208, USA

[2] Department of Counseling, School and Educational Psychology, University at Buffalo, The State University of New York, Buffalo, NY 14260-1000, USA

symptoms of children with ASD/HFASD. Stratis and Lecavalier (2015) conducted a meta-analysis of informant agreement for rating scales measuring emotional/behavioral symptoms and social skills of youth with ASD and found low-to-moderate agreement between different raters ($r = .28$–$.39$; e.g., parent-teacher, parent–child). Further, IQ and age (as continuous variables) moderated agreement among some raters for internalizing symptoms. Although their study provided valuable information about informant discrepancies, it was not focused specifically on ASD symptoms and most of the studies relied on Pearson correlations. Stratis and Lecavalier (2015) recommended further studies of parent-teacher agreement, as well as an expanded examination of moderators.

The Developmental Social Disorders (DSD) content scale of the Behavior Assessment System for Children-Second Edition (BASC-2; Reynolds and Kamphaus 2004) measures ASD-related symptoms involving social-communication and narrow and repetitive interests and activities. The DSD scale is of interest as the BASC-2 is a widely used and broad-based initial screening measure. Parents and teachers can complete the rating scale, which allows for cross-informant comparisons. No data are presented in the manual on parent-teacher inter-rater reliability for the DSD scale; however moderate correlations are reported between their ratings for the other BASC-2 clinical scales ($M$ $r$s = .38–.39). Volker et al. (2010) found that DSD parent ratings effectively discriminated children with HFASD from typically-developing children (no teacher ratings were reported). Despite its potential utility, no studies have comprehensively examined parent-teacher agreement and reliability, and moderators of discrepancies of the DSD scores for children with HFASD.

There is a need for comprehensive examination of informant discrepancies for scales assessing ASD-related symptoms for children with HFASD and available data are largely restricted to examination of correlations (Pearson $r$). According to Stolarova et al. (2014), studies of cross-informant ratings should be more comprehensive and include intra-class correlations coefficients (ICCs), as well as simple linear correlations (Pearson). Between-group differences in mean scores and the clinical ranges of the children should also be considered (Achenbach 2011). Lastly, informant discrepancy studies should utilize standard scores in order to increase the interpretability of findings across studies (Achenbach 2011; De Los Reyes and Kazdin 2004). This study comprehensively analyzed informant discrepancies for parent and teacher ratings of ASD-related symptoms/problems using the DSD for children with HFASD. The study included between-group mean and clinical range comparisons, ICC and Pearson correlations, a Bland–Altman plot, and moderator analyses.

## Method

### Participants

The sample consisted of two informant groups from which 240 ratings were collected using the DSD scale ($n = 120$ parents/caregivers and $n = 120$ teachers). Parents and teachers provided ratings of the same 120 children, ages 6–12 years with HFASD. The large majority of parent raters were mothers ($n = 117$; 97.5 %). For the teacher raters, the mean number of days teaching the children in the sample was 85 days (range from 28 to 195 days). The teacher raters did not report whether they were special education or general education teachers, however 82.5 % ($n = 99$) of the children were receiving special education services. The children were participants from several clinical trials evaluating psychosocial treatments for children with HFASD. These treatment trials were testing the efficacy of various psychosocial interventions targeting social/social-communication skills and ASD symptoms. As previously noted, children with HFASD are characterized by social-communication impairments and circumscribed and repetitive behaviors and interests, but exhibit relatively intact cognitive and language abilities (APA 2013). All children had a prior clinical diagnosis of autism, Asperger's, or PDD-NOS, Wechsler Intelligence Scale for Children-4th Edition (Wechsler 2003) short-form IQ > 70, and Comprehensive Assessment of Spoken Language (Carrow-Woolfolk 1999) short-form Expressive or Receptive language score >70. In addition, all met criteria on the Autism Diagnostic Interview-Revised (Rutter et al. 2003), which was completed to verify diagnosis. The child sample was mainly male (90.0 %) and Caucasian (97.5 %), with a mean IQ and language level in the average range (Table 1). Average parent education was 13.8 years.

### Measure

*Behavior Assessment System for Children-Second Edition, Parent Rating Scales (BASC-2 PRS) and Teacher Rating Scales (BASC-2 TRS)*

The BASC-2 PRS and TRS (Reynolds and Kamphaus 2004) assesses symptoms and behaviors across a wide variety of domains using a child form (ages 6–11; PRS-C and TRS-C) and an adolescent form (ages 12–21; PRS-A and TRS-A). The symptoms and behaviors assessed range from externalizing problems (e.g., Conduct Problems, Hyperactivity, etc.), to internalizing problems (e.g., Anxiety, Depression, etc.), to adaptive skills (e.g., Adaptability, Social Skills, etc.). The "scales are consistent not only across sex and age levels but also between the teacher and

**Table 1** Demographic characteristics of child sample

| Characteristic | Child participants ($n = 120$) | |
|---|---|---|
| | *M* (*SD*) | |
| Age in years | 8.78 (1.36) | |
| WISC-IV short-form IQ | 103.16 (14.35) | |
| CASL | | |
| Short-form expressive language | 98.29 (15.48) | |
| Short-form receptive language | 101.41 (16.81) | |
| ADI-R | | |
| Impairment in social interaction | 19.12 (5.80) | |
| Impairment in communication | 15.51 (4.82) | |
| Restricted repetitive behavior | 5.85 (2.15) | |
| | *n* (% of total) | |
| Gender | | |
| Male | 108 (90) | |
| Female | 12 (10) | |
| Ethnicity | | |
| Caucasian | 117 (97.5) | |
| Latino | 2 (1.7) | |
| Other | 1 (.8) | |
| BASC-2 composites | Parent ratings | Teacher ratings |
| | *M* (*SD*) | *M* (*SD*) |
| Internalizing composite | 55.28 (11.90) | 57.20 (12.26) |
| Externalizing composite | 56.72 (10.83) | 54.21 (9.63) |
| Adaptive skills composite | 34.57 (8.56) | 39.97 (5.70) |

WISC-IV 4-subtest short-form consisted of the block design, similarities, vocabulary, and matrix reasoning subtests; CASL 4-subtest short-form consisted of the antonyms, synonyms, syntax construction, and paragraph comprehension subtests. The WISC-IV and CASL scores are standard scores with a normative $M = 100$ and $SD = 15$. The BASC-2 composite scores are standard scores with a normative $M = 50$ and $SD = 10$. The age range of the child participants was 6–12 years

*WISC-IV* wechsler intelligence scale for children-4th edition, *CASL* comprehensive assessment of spoken language, *ADI-R* autism diagnostic interview-revised, *BASC-2* behavior assessment system for children, second edition

parent forms … [providing] a basis for consistent interpretation of scales and for meaningful across-source … score comparisons" (Reynolds and Kamphaus 2004, p. 2). The PRS and TRS forms also include the Developmental Social Disorders (DSD) content scale. Items on the DSD scale (across the age and source forms) assess ASD-related symptoms involving deficits in social and communication skills (e.g., easily makes friends, has difficulty making friends, communicates clearly with others, etc.), and interests and activities (e.g., exhibits interest in the ideas of others, is able to adjust to changes in routine, etc.). The DSD scale is comprised of items from the other clinical scales that were configured to specifically assess ASD-related symptoms. The DSD scale on the PRS-C, TRS-C, and TRS-A forms each included 14 items and the PRS-A includes 12 items (in this study, only one child was 12 years of age and was rated using the PRS-A and TRS-A forms). Parents and teachers rate items on a scale from 0 (never) to 3 (almost always). DSD item scores are summed and converted to a standardized *T*-score ($M = 50$, $SD = 10$) which is indicative of severity (60–69 at-risk and $\geq$70 clinically significant). The continuous scaling of the BASC-2 makes it useful as a measure of treatment responsiveness. Internal consistency estimates of the DSD scale ranged from .82–.85 (PRS-C and -A) and .89–.90 (TRS-C and -A). Adjusted test–retest reliabilities were .82 and .78 for the PRS-C and -A, respectively, and .90 and .88 for the TRS-C and -A, respectively. Moderate correlations were reported between the DSD scale and subscales of other measures assessing related symptoms/behaviors (correlations generally in the .40–.60 range for scales such as withdrawal, shyness, social skills). Evidence in the test manual indicates that the DSD scale accurately discriminates youth with ASD from non-ASD youth (see Reynolds and Kamphaus 2004).

## Procedures

The treatment trials from which the cases were drawn were approved by the Institutional Review Board and conducted in compliance with the approved procedures. For those treatment trials, parents and teachers completed the entire BASC-2 (which included the DSD scale) as one of the baseline (pretest) measures; each child was rated by one parent and one teacher. Each form was checked upon return for completeness and any errors (multiple endorsements, omitted items, etc.) were immediately corrected with the informant. This resulted in all protocols being complete and no missing items. All protocols were scored by graduate-level research assistants using the BASC-2 ASSIST Plus computer scoring software which includes a second entry check for accuracy. The DSD scores (and other data) were entered into a database and independently checked by a second member of the team.

## Data Analysis Plan

In the current study, the sample size of 120 pairs of raters was sufficient to detect a correlation of .30 at an alpha of .05 with power of .79 (ESCI, Power-Paired; Cumming, 2012). The correlation of .30 was identified as the minimally important value based on prior informant discrepancy reliability studies of youth with ASD (Schanding et al. 2012; Stratis and Lecavalier 2015). Data analyses included (a) sample means versus estimated population means from the test manual (one-sample *t*-tests), (b) parent versus

teacher means (paired *t*-test), (c) classification accuracy of parent and teacher scores (percentages at cutpoints corresponding to clinical severity categories, percent agreement by clinical severity category), (d) inter-rater agreement and consistency (ICC, Pearson correlation, Bland–Altman plot, and regression), and (e) examination of potential moderators of parent-teacher discrepancies (Pearson correlations and scatterplots). Effect sizes and confidence intervals were calculated for all estimates. Bland–Altman plots examine the degree of agreement between two sets of raters or scales, in this case agreement of parents and teachers. They also determine whether there are any systematic trends across the ranges of scores (Bland and Altman 1986). The vertical (y) axis is the difference between parent and teacher ratings; the horizontal (x) axis is the mean of the parent-teacher pairs of ratings (Fig. 1). The solid horizontal line references the mean difference score, and the dotted lines above and below represent the 95 % confidence interval for the differences. The teacher score was subtracted from the parent score; thus, a positive difference score indicates that the parent submitted a higher rating than the teacher for the particular child. Bland and Altman (1986) recommended regression analysis to examine the possibility of systematic trends across the scale range by regressing the mean onto the difference score.

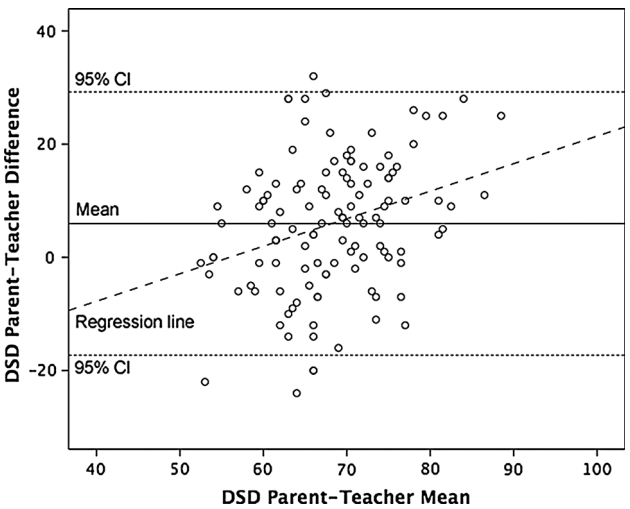


**Fig. 1** Bland-Altman plot of the BASC-2 DSD. The plotted variables (parent-teacher difference scores and means) were normally distributed. The *vertical* (*y*) *axis* is the difference between parent and teacher ratings; the *horizontal* (*x*) *axis* is the mean of the parent-teacher pairs of ratings. The *solid horizontal line* references the mean difference score, and the *dotted lines* above and below represent the 95 % confidence interval for the differences. A positive difference score indicates that the parent submitted a higher rating than the teacher for the particular child

## Results

### Normative and Cross-Informant Comparisons

Initially, data for the individual informant groups were compared to population estimates. Both the mean DSD parent ratings, $t$ (119) = 21.71, $p < .001$, $d = 2.13$ (large effect), and teacher ratings, $t$ (119) = 21.12, $p < .001$, $d = 1.53$ (large effect), were significantly different than the population estimate. Specifically, the mean parent and teacher ratings were significantly higher than the estimated population mean. The cross-informant group (parent vs. teacher) comparison was then conducted (Table 2). Results indicated a significant between-groups difference ($p < .001$; medium effect size $d = .63$), with the mean DSD parent score falling in the clinically significant range ($M = 71.26$) and teacher score in the at-risk range ($M = 65.28$).

Data were also examined to determine the percentage of children in the sample scoring above prescribed clinically-important cutpoints. For parents and teachers respectively, the percentage of DSD scores falling $\geq$60 (at-risk) was 83 and 79 %, and $\geq$70 (clinically significant) was 60 and 29 %. Parent-teacher agreement by clinical severity category was also assessed by calculating percent agreement for dichotomously coded non-clinical (within normal limits <60) versus clinically important ($\geq$60) categories, as well as for each clinical severity sub-category. Dichotomously-coded parent and teacher DSD ratings would place the child in the same category in 87 of the 120 cases (73 %) (clinically-important = 81/100, 81 %; within normal limits = 6/20, 30 %). Categorical agreement by severity subcategories (Table 3) was 38 % overall, with the highest level of agreement in the at-risk category (52 %).

### Parent-Teacher Reliability Estimates

Two coefficients were calculated in order to examine parent-teacher reliability in terms of agreement and linear relationship (Table 2). The coefficients were statistically significant but reflected generally low agreement (ICC = .30 and Pearson = .22). A Bland–Altman plot of the DSD scores was created to further examine the degree of agreement between parents and teachers, and to determine whether there was any systematic trend across the range of scores (Fig. 1). The mean parent-teacher difference for the DSD scale was 5.98 ($SD = 11.86$), with the largest differences being +32 and −24. A positive difference score indicates that the parent provided a higher rating than the teacher for the particular child. The DSD parent-teacher means averaged 68.27 ($SD = 7.33$). Bland and Altman (1986) recommended regression analysis to examine the possibility of systematic trends across the

**Table 2** Mean difference and correlations between parent and teacher ratings on the BASC-2 DSD

| Parent | Teacher | Mean score comparison | | | ICC | | Pearson *r* | |
|---|---|---|---|---|---|---|---|---|
| *M* (*SD*) | *M* (*SD*) | *t*(119) | *p* | *d* [95 % CI] | Coefficient (*p*) | [95 % CI] | Coefficient (*p*) | [95 % CI] |
| 71.26 (10.73) | 65.28 (7.92) | 5.53 | <.001 | .63 [.39, .87] | .30 (.011) | [.01, .50] | .22 (.017) | [.04, .38] |

All calculations based on parent and teacher ratings of $n = 120$ children with HFASD. Standards for Cohen's *d*: small = .20, medium = .50, and large = .80 (Cohen 1988). ICC indicates the variance of ratings of an individual child by two informants and the variance across the entire sample of children. ICC calculated as two-way mixed effects model using absolute agreement definition of reliability (Shrout and Fleiss 1979). Pearson *r* indicates the linear correlation between the two informant groups

**Table 3** Parent and teacher BASC-2 DSD agreement by clinical category

| | Teacher | | | % Agreement |
|---|---|---|---|---|
| | Non-clinical | At-risk | Clinically significant | |
| *Parent* | | | | |
| Non-clinical ($n = 20$) | **6** | 9 | 5 | 30 |
| At-risk ($n = 29$) | 9 | **15** | 5 | 52 |
| Clinically significant ($n = 71$) | 10 | 36 | **25** | 35 |
| Overall | | | | 38 |

Bold values indicate the number of cases in which the parent and teacher both rated the child in the same clinical severity category

BASC-2 DSD categories defined in the test manual as: Non-clinical <60, At-risk 60–69, and Clinically Significant ≥70

scale range by regressing the mean onto the difference score. The regression analysis showed a significant positive relationship between the difference scores and the means ($B = .486$, $t = 3.42$, $p = .001$), which is depicted in the plot. This indicated that the discrepancy between the parent–teacher ratings appeared to increase as the mean of their ratings increased.

### Analysis of Potential Moderators of Parent-Teacher Discrepancies

To explore whether individual difference characteristics of children, parents, and/or teachers might moderate parent-teacher difference scores, scatterplots and correlations were examined. Child variables included age, WISC-IV (short-form) IQ, and CASL (short-form) Expressive and Receptive language scores. Parent average education and teacher familiarity with the child (i.e., number of days from start of school year until assessment) were also studied. There were no significant correlations between any potential moderators and parent-teacher difference scores, and scatterplots provided no evidence of nonlinear relationships.

## Discussion

Little empirical research has comprehensively examined informant discrepancies for scales assessing ASD-related symptoms for children with HFASD, and no studies were identified that have done so using the BASC-2 DSD. Such studies are needed as rating scales including the BASC-2 are commonly used by clinical practitioners, as well as by researchers studying children with HFASD (e.g., Volker et al. 2010). This study comprehensively examined parent-teacher discrepancies in DSD ratings for children with HFASD. Mean DSD scores from both informant groups were significantly higher than the population estimates. Findings of significant elevations in parent and/or teacher DSD scores for children with HFASD are consistent with prior findings comparing children with HFASD to typically-developing children (Volker et al. 2010), as well as with data in the BASC-2 test manual (Reynolds and Kamphaus 2004) indicating that the DSD is sensitive to ASD-related symptoms. Results of the current study also found that parent ratings were significantly higher than teacher ratings, with a mean difference of approximately 6 *T*-score points (medium effect size). Despite this difference, parents and teachers agreed on 81 % of the cases (children) having symptoms at or above the minimal cut-score (60) and 95 % of the cases had at least one of the informants rating the child above the clinical threshold. Only 6 cases (5 % of the total sample) had both informants' ratings below the clinical threshold indicating a low likelihood of both informants failing to perceive clinically important ASD-related symptoms.

Parent-teacher ratings were also analyzed for their strength of association. Coefficients were statistically significant but low in magnitude for both agreement (ICC) and linear relationship (Pearson) between the informants' scores. Regression analysis indicated a significant trend

between the difference scores and the means. As reflected in Fig. 1, the difference between the parent-teacher ratings appeared to increase as the mean of their ratings increased. Specifically, the magnitude of the difference (parent > teacher) increased as the informants perceived severity of ASD-related symptoms increased. This suggests that clinicians may encounter more significant discrepancies in DSD scores when parents and teachers perceive high levels of ASD symptoms. This pattern was also supported in the higher percentage of agreement when parents and teachers rated the children's symptoms in the less severe *at-risk* category.

Taken together, the pattern of scores suggests that clinicians will likely encounter significant discrepancies between parent and teacher DSD scores when gathering information on children with HFASD. The scores will be positively correlated however will likely become more discrepant, on average, as parents and teachers perceive the symptoms as being more severe. Given the critical importance of detection during broad-based initial screening, a conservative approach would be to consider a score above the minimum threshold ($T \geq 60$) from a parent or teacher as clinically important and sufficient for referral for a more comprehensive ASD assessment. Since the DSD score can also be used as part of a comprehensive assessment to inform a diagnostic decision and/or as a measure of ASD-related symptom severity in response to a treatment (Reynolds and Kamphaus 2004), clinicians will need to gather additional information from informants when discrepancies occur. Because rating scales are not diagnostic on their own (Norris and Lecavalier 2010) clinicians should consider all sources of data along with the DSD score to inform their diagnostic decisions. Clinician observations can also assist in clarifying parents and teachers perceptions and the discrepancy. These efforts will help determine the reason(s) for the disagreement (e.g., informants' perceptions of the symptoms, understanding of the items, and/or real variability in the symptoms across settings; Achenbach 2011; De Los Reyes 2011). Results also suggested that the individual child, parent, and teacher variables examined in this study will not likely moderate the parent-teacher discrepancies in DSD scores for children with HFASD.

This study had a number of strengths including a comprehensive examination of parent-teacher discrepancies, evaluation of a scale from a commonly used broad-based rating scale, and use of standard scores in the analyses. It was also unique in the use of a Bland–Altman plot and examination of moderators of the discrepancies for the DSD scale. Finally, this study met the need for rigorous studies of rating scales using a well-characterized sample of children with HFASD (Norris and Lecavalier 2010). Despite these strengths, several limitations warrant mention. The sample was largely male and Caucasian and was high-functioning which limits generalizability. Although there is a need to study ratings in more functionally-homogeneous ASD sub-groups (as ratings and discrepancies can vary as a function of developmental level; Achenbach 2011), the inclusion criteria for the child sample may have restricted the variability for the child factors used in the moderator analyses. Lastly, this study only examined parent-teacher ratings and comprehensive studies are needed of discrepancies for other informants (e.g., parent–parent, parent–child) and/or other rating scales for children with HFASD.

**Funding** The research reported in this article was supported by Department of Education, Institute of Education Sciences Grants R324A130216 and R324A080136, and a research grant from the Organization for Autism Research (OAR). Findings and conclusions are those of the authors and do not necessarily reflect the views of the funding agencies.

**Author Contributions** CL conceived of the study, participated in its design, and drafted the manuscript; JPD participated in the design, conducted the statistical analyses, and assisted in the interpretation of the data and preparation of the manuscript; AKJ participated in the study design and coordination, and assisted in manuscript preparation; MLT led the data collection and participated in the study design and manuscript preparation; CAM participated in the study design, assisted with data management and coordination, and contributed to manuscript preparation; and JDR assisted with data management and coordination and contributed to manuscript preparation. All authors read and approved the final manuscript.

**Compliance with Ethical Standards**

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

**Ethical Approval** All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed Consent** Informed consent was obtained from all participants included in this study.

## References

Achenbach, T. M. (2011). Commentary: Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *Journal of Child and Adolescent Psychology, 40*(1), 80–86. doi:10.1080/15374416.2011.533416.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213–232. doi:10.1037/0033-2909.101.2.213.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington: American Psychiatric Association.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 8476*, 307–310.

Carrow-Woolfolk, E. (1999). *Comprehensive assessment of spoken language*. Circle Pines: American Guidance Service.

Center for Disease Control and Prevention (CDC). (2014). Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 states, United States, 2010. *MMWR, 63*(SS-2), 1–21.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum.

Constantino, J. N., & Gruber, C. P. (2012). *Social responsiveness scale, second edition (SRS-2)*. Torrance: Western Psychological Services.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 40*(1), 1–8. doi:10.1080/15374416.2011.533405.

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment, 16*(3), 330–334. doi:10.1037/1040-3590.16.3.330.

Norris, M., & Lecavalier, L. (2010). Screening accuracy of level 2 autism spectrum disorder rating scales: A review of selected instruments. *Autism, 14*(4), 263–284. doi:10.1177/1362361309348071.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior assessment system for children* (2nd ed.). Circle Pines: AGS.

Rutter, M., LeCouteur, A., & Lord, C. (2003). *Autism diagnostic interview-revised*. Los Angeles: Western Psychological Services.

Schanding, G. T., Nowell, K. P., & Goin-Kochel, R. P. (2012). Utility of the social communication questionnaire-current and social responsiveness scale as teacher-report screening tools for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*, 1705–1716. doi:10.1007/s10803-011-1412-9.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428. doi:10.1037/0033-2909.86.2.420.

Stolarova, M., Wolf, C., Rinker, T., & Brielmann, A. (2014). How to assess and compare inter-rater reliability, agreement and correlation of ratings: An exemplary analysis of mother-father and parent-teacher expressive vocabulary rating pairs. *Frontiers in Psychology, 5*, 1–13. doi:10.3389/fpsyg.2014.00509.

Stratis, E. A., & Lecavalier, L. (2015). Informant agreement for youth with autism spectrum disorder or intellectual disability: A meta-analysis. *Journal of Autism and Developmental Disorders, 45*, 1026–1041. doi:10.1007/s10803-014-2258-8.

Volker, M. A., Lopata, C., Smerbeck, A. M., Knoll, V., Thomeer, M. L., Toomey, J. A., & Rodgers, J. D. (2010). BASC-2 PRS profiles for students with high-functioning autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*, 188–199. doi:10.1007/s10803-009-0849-6.

Wechsler, D. (2003). *Wechsler intelligence scale for children* (4th ed.). San Antonio: The Psychological Corporation.