

Published in final edited form as:

J Autism Dev Disord. 2014 October ; 44(10): 2400–2412. doi:10.1007/s10803-012-1719-1.

Standardizing ADOS Domain Scores: Separating Severity of Social Affect and Restricted and Repetitive Behaviors

Vanessa Hus,

Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109, USA
vhus@umich.edu

Katherine Gotham, and

Vanderbilt Kennedy Center, Vanderbilt University, Nashville, TN, USA

Catherine Lord

Center for Autism and the Developing Brain, Weill Cornell Medical College, White Plains, NY, USA

Abstract

Standardized Autism Diagnostic Observation Schedule (ADOS) scores provide a measure of autism severity that is less influenced by child characteristics than raw totals (Gotham et al. in *Journal of Autism and Developmental Disorders*, 39(5), 693–705 2009). However, these scores combine symptoms from the Social Affect (SA) and Restricted and Repetitive Behaviors (RRB) domains. Separate calibrations of each domain would provide a clearer picture of ASD dimensions. The current study separately calibrated raw totals from the ADOS SA and RRB domains. Standardized domain scores were less influenced by child characteristics than raw domain totals, thereby increasing their utility as indicators of Social-Communication and Repetitive Behavior severity. Calibrated domain scores should facilitate efforts to examine trajectories of ASD symptoms and links between neurobiological and behavioral dimensions.

Keywords

Autism spectrum disorders; Autism Diagnostic Observation Schedule; Severity; Social Affect; Restricted and Repetitive Behaviors

Introduction

The search to elucidate underlying biological mechanisms which cause or increase risk for autism spectrum disorders (ASD) has been made more complicated by the marked phenotypic heterogeneity associated with this developmental disorder (State and Levitt

© Springer Science+Business Media New York 2012

At the start of this study, Vanessa Hus, Katherine Gotham, and Catherine Lord were at the University of Michigan, Department of Psychology and University of Michigan Autism & Communication Disorders Center in Ann Arbor, Michigan.

Vanessa Hus has remained at the University of Michigan, Department of Psychology. Katherine Gotham is now at the the Vanderbilt Kennedy Center, Nashville, Tennessee. Catherine Lord is now at the Center for Autism and the Developing Brain, Weill Cornell Medical College, White Plains, New York.

Conflict of interest C. Lord receives royalties for the ADOS; profits from this study were donated to charity.

2011). Diagnostic criteria focus on the presence or absence of specific behaviors or impairments in three domains: *Communication*, *Reciprocal Social Interaction*, and *Restricted and Repetitive Stereotyped Behaviors and Interests* (American Psychiatric Association 2000; World Health Organization [WHO] 1992). However, ASD symptoms within each domain vary considerably in type and severity, depending upon an individual's age, language level, and IQ.

Current nosology attempts to capture some of this variation through categorical diagnoses (e.g., Autistic Disorder, Asperger's Disorder and Pervasive Developmental Disorder, Not Otherwise Specified; APA 2000). However, research has demonstrated that differentiations made between ASD subgroups are often not reliable across different sites (Lord et al. 2011). In addition, in several studies, items reflecting social and communication impairments comprised a single factor on ASD diagnostic instruments (e.g., Frazier et al. 2012; Gotham et al. 2007). In light of these findings, proposals for DSM-5 and ICD-11 call for subgroups to be subsumed into a single category of ASD defined by two behavioral domains: *Social/Communication Deficits* and *Fixated or Restricted Interests and Repetitive Behaviors* (APA 2011, WHO 2012). Several initial studies support these proposed changes (Frazier et al. 2012; Huerta et al. in press, Mandy et al. 2012, though see Mattila et al. 2011 and McPartland et al. 2012). To further capture the heterogeneity, criteria for assessing severity within each domain are recommended.

As these changes are implemented, many of the currently used ASD diagnostic instruments will need to be revised to more accurately reflect new DSM-5 and ICD-11 criteria, both to inform diagnosis and to describe severity of symptoms within each behavioral domain. For example, the diagnostic algorithm of the Autism Diagnostic Interview—Revised (ADI-R; Rutter et al. 2003), a widely-used parent interview in autism research, is divided into three domains reflecting the current DSM-IV and ICD-10 criteria for Autistic Disorder, whereas the Social Responsiveness Scale (SRS; Constantino and Gruber 2005), a caregiver questionnaire, relies on a single total score for diagnostic classification. In contrast, the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al. 2012), a clinician-administered observational assessment, has recently revised diagnostic algorithms that comprise two behavioral domains [referred to as Social Affect (SA) and Restricted and Repetitive Behaviors (RRB)] and provide cut-offs for ASD classification (Gotham et al. 2007). In addition, total scores from the revised ADOS algorithms have been standardized to provide a continuous measure of overall ASD symptom severity that is less influenced by child characteristics, such as age and language skills, than raw totals [Calibrated Severity Scores (CSS); Gotham et al. 2009]. These scores can be used to compare ASD symptom severity across individuals of different developmental levels. As such, they provide a “purer” metric of overall ASD severity than raw totals from the ADI-R and SRS, for which studies have demonstrated strong influences of child characteristics, such as age, language level, and non-ASD specific behavior problems (e.g., Constantino et al. 2003; Hus et al. in press; Hus and Lord in press).

Although the ADOS-CSS may provide some advantages over these other measures of general ASD severity, the nature of the symptoms underlying an individual's CSS may vary greatly. For example, an ADOS-CSS of 10, indicating the highest level of severity, may be

assigned to a child with very significant social-communication impairments who exhibits few repetitive behaviors during the ADOS. The same score may also be assigned to a child who has moderate levels of impairments in both domains or very high levels of repetitive behaviors and more subtle social-communication difficulties. Social-communication difficulties often pertain to a “lack” of typical behaviors that are pervasive across social contexts, such as reduced use of gestures or eye contact or reduced frequency of appropriate social responses, making them more easily-observable during brief interactions. In comparison, RRBs are often characterized by the presence of an abnormal behavior, such as hand flapping, sensory examination of materials or excessive references to a particular topic. Because RRBs may only occur in particular conditions (e.g., hand flapping when a child is very excited or prolonged discussion of a topic only if it is raised), it is more difficult to assess them in a short period of time. Therefore, it is important to acknowledge that, when assessing and comparing symptom severity in different domains, the ADOS as a source of information, particularly about RRBs, is limited by both time and context. While the presence of RRBs during this brief observation may be clinically significant, the absence of these behaviors in this time-limited, standardized context must be interpreted more cautiously. Nevertheless, research has suggested that both social-communication and repetitive behaviors measured by the ADOS are surprisingly good predictors of diagnosis (e.g., Lord et al. 2006).

Separate calibration of these distinct domains is needed to provide a clearer picture of ASD severity. For example, calibrated domain scores would allow for examination of two dimensions (SA and RRB), which may have distinct developmental trajectories or respond differently to intervention. In large samples, researchers could use estimates of social-communication and repetitive behavior severity to increase phenotypic homogeneity by clustering individuals according to similar levels of severity in each domain (e.g., high SA and RRB; high SA and low RRB, etc.). In smaller studies that cannot afford the loss of power resulting from sample stratification, researchers might use continuous scores to statistically control for differences in one domain while focusing on the other. Separately calibrated domain scores may also be useful in genetic and neurobiological studies seeking to draw associations between biological mechanisms and specific behavioral domains, many of which currently rely on raw domain totals (e.g., Dichter et al. 2011). While some studies have controlled for effects of age or IQ in individual samples (e.g., Di Martino et al. 2011), use of calibrated scores may facilitate comparisons across samples comprised of individuals of varying developmental levels.

The goal of the current study was to separately calibrate raw totals from the ADOS SA and RRB domains to reduce the effects of child characteristics and increase the utility of these scores as continuous measures of social-communication and of repetitive behavior symptom severity.

Methods

Participants

For comparability, the same sample used to standardize the overall ADOS total (see Gotham et al. 2009) was also employed to calibrate separate severity metrics for the Social Affect

(SA) and Restricted, Repetitive Behavior (RRB) domains. Briefly, this included data from 1,415 individuals ranging in age from 2 to 16 years. With repeated assessments for 25 % of the sample, data from 2,195 ADOSes with contemporaneous best estimate clinical diagnoses were available for analysis. Of these assessments, 1,786 cases were given an autism spectrum disorder diagnosis (ASD; 1,187 Autistic Disorder, 599 Other-ASD) and 409 had a Non-ASD diagnosis. Non-ASD diagnoses included language disorders (27 %), nonspecific intellectual disability (20 %), Down syndrome (14 %), oppositional defiant disorder or ADD/ADHD (13 %), mood or anxiety disorders (8 %), Fetal Alcohol Spectrum Disorders (7 %), other genetic or physical disabilities, such as Fragile X or mild cerebral palsy (6 %) and early developmental delays (5 %).

Individuals were consecutive referrals to specialty clinics in Ann Arbor, Michigan and Chicago, Illinois, and participants in research studies conducted through the University of North Carolina—Chapel Hill, University of Chicago, and University of Michigan. All participants provided informed consent and all procedures related to this project were approved by institutional review boards at the University of Chicago or University of Michigan. Sample characteristics are provided in Table 1.

Procedure

The ADOS was conducted as part of a clinical or research evaluation (see Gotham et al. 2009 for more detailed procedures). All ADOSes were administered and scored by a clinical psychologist or trainee who met standard requirements for research reliability. The Pre-Linguistic ADOS (PL-ADOS; DiLavore et al. 1995) was given in 418 (19 %) assessments and a pilot version of the ADOS-Toddler (Luyster et al. 2009) was given in 82 assessments (4 %). For both measures, scores from items identical to those in the Module 1 algorithms were used. Verbal and/or nonverbal IQ scores were available for 2009 (92 %) assessments. These were derived from a developmental hierarchy of cognitive measures (see Lord et al. 2006), most frequently the Mullen Scales of Early Learning (Mullen 1995) and the Differential Ability Scales (Elliott 1990). Best estimate clinical diagnoses were made by a supervising clinical psychologist and/or a child psychiatrist after review of all assessment data (including, at a minimum, the ADOS and cognitive scores).

Standardization of Raw Totals

Calibration of each domain began by following a similar procedure to that described for standardization of overall ADOS totals (Gotham et al. 2009). Only assessments from individuals with ASD were used for raw domain total standardization. This included all assessments with a corresponding best estimate clinical diagnosis of autism or Other-ASD, as well as data from 13 individuals who had ADOS data with a contemporaneous Non-ASD diagnosis but who were later diagnosed with ASD (total $n = 1,807$ assessments from 1,118 individuals). Participants were first divided into the 18 age/language groups used for the calibration of the overall raw totals. SA and RRB scores were compared separately for each 1-year chronological age group within a given cell to ensure that distributions of the domain scores were comparable. Some of the 18 cells were then collapsed due to comparable distributions (likely due to the reduced range of scores in each domain compared to the overall totals). This resulted in 12 age/language cells (See Fig. 1; note that the raw total-to-

calibrated score mapping for the RRB domain could have been further collapsed into two Module 2 cells, 2–3 year olds and above 4 years; however, these were left expanded across 4 cells so that both domains would have the same number of cells).

In the overall total calibration, ADOS diagnostic classifications were used to anchor raw totals to ranges of severity scores. That is, raw totals corresponding to an ADOS classification of “Autism” were mapped on to CSS of 6–10, “ASD” to CSS of 4–5 and “Nonspectrum” to CSS of 1–3. This was done to make the metric more generalizable to other samples, as we cannot assume that the datasets used for calibration in all developmental cells were representative of the heterogeneous ASD population. Next, the range of raw totals assigned to each point on the 10-point severity scale was determined by the percentiles of available data within that classification range (Gotham et al. 2009). Because there are not separate SA cut-offs for “Autism” and “ASD” classifications, the same percentiles used for mapping raw ADOS totals (i.e. SA + RRB) to the 10-point scale were used to inform the mapping of raw SA totals to SA-CSS within each of the 12 age/language cells. Raw total-to-calibrated score mappings were then adjusted so that, for each of the 5 diagnostic algorithm groups (Gotham et al. 2007), sensitivity for individuals receiving an ADOS classification of “Autism” and an SA-CSS greater than or equal to 6 was, if possible, at or above 90 %. Within algorithm groups, the lowest individual cell sensitivity was .89 for Module 2, 2–3 year olds. A goal of 80 % sensitivity across algorithm groups was set for individuals with an “Autism Spectrum” ADOS classification and an SA-CSS of 4 or higher. Sensitivity for individual developmental cells within algorithm groups was sometimes lower in groups with few participants; however, considering cells with greater than 20 participants, only Module 3, 3–5 year olds ($n = 59$) fell just below this threshold, with a sensitivity of .78. Finally, adjustments were made to ensure that specificity (individuals with a “Nonspectrum” ADOS classification and SA-CSS less than or equal to 3) was, if possible, at least 80 % for each algorithm. Within algorithms, only the Module 2, 5–6 year old cell fell below this threshold, with a specificity of .76.

Because the RRB domain is limited to a range of 9 points (0–8), it was not possible to use all 10 points in the severity metric for this domain. However, given concerns that SA- and RRB-CSS scores may be misinterpreted if they are not on a comparable scale, it was decided to maintain the full 10-point range and have some points on the severity scale for which no raw scores were assigned. Thus, as with SA-CSS, percentiles from mapping of raw overall totals were used to inform mapping of raw RRB totals to the calibrated metric. This resulted in the raw RRB totals mapping on to CSS values of 5–10. These distributions were skewed compared to the overall and SA-CSS scales and reflect the trade-off in using the ADOS as a measure of RRBs: while a lack of RRBs is difficult to interpret, the presence of RRBs during this brief observation is more meaningful as an indication of greater severity. Given the lower sensitivity of repetitive behaviors in the limited context in which they may be observed during the ADOS, a goal of 80 % sensitivity was set for individuals receiving an ADOS classification of “Autism” and RRB calibrated scores of 6 or greater; Module 3, 2–5 year olds fell just below this threshold with a sensitivity of 77 %. No sensitivity threshold was set for individuals with an “Autism Spectrum” classification. A goal of 80 % specificity was set for scores less than or equal to 6. For individual cells with greater than 20 participants, the lowest specificity was 79 % for Module 3, 6–16 year olds.

Table 2 shows the mappings of raw SA and RRB totals to the 10-point severity scale for each of the 12 calibration cells.

Associations Between Participant Characteristics, Raw Domain Totals and Calibrated Domain Scores

Following procedures in Gotham et al. 2009, separate linear regression analyses were conducted using the sample of participants with ASD who had contemporaneous demographic data ($N = 1,369$) to examine the influences of child characteristics on raw domain totals and calibrated domain scores. The child's verbal and nonverbal IQs and mental ages were entered into the first block, followed by child chronological age, gender, maternal education and race in the second block. Only model R^2 are reported because interpretation of the meaning of these individual coefficients is limited by multicollinearity. Next, significant predictors were entered into Forward Stepwise models to assess the relative contributions of these variables in predicting raw domain totals and calibrated domain scores. (Results from analyses including Non-ASD participants are available from authors. Consistent with the results for the participants with ASD, when applied to the entire clinically-referred sample, standardized severity scores were less influenced by participant characteristics than were raw domain totals.)

Results

Comparison of Raw Domain Totals and Calibrated Domain Scores by Calibration Cell

As shown in Table 3 and Fig. 2a, c, distributions of raw SA and RRB domain totals varied significantly by age/language group. Across algorithms reflecting different language levels, individuals with less language had higher scores than those who were more verbally fluent. Within algorithm groups, older children and adolescents tended to have higher scores than toddlers and young children. In contrast, calibrated SA and RRB domain scores were more comparable across calibration cells, though not uniform (see Table 3 and Fig. 2c, d). Notably, children who were verbally fluent (i.e., Module 3) have a wider distribution of RRB-CSS scores compared to children of other language levels. This reflects the somewhat larger proportion of verbally fluent children (8.5–12.9 %) that did not have repetitive behaviors during the ADOS (i.e., received a RRB-CSS of 1).

As noted above, ADOS *classifications*, which are based on raw overall totals (SA + RRB) were used to anchor the raw total-to-overall severity score mappings for the domains to specific calibrated score ranges (e.g., "Autism" to CSS of 6–10). Using percentiles from the raw total-to-overall CSS mapping to inform raw domain totals-to-domain severity score mappings, mean SA-CSS and RRB-CSS also distinguished between individuals grouped by clinicians' best estimate clinical diagnoses (i.e., Autism vs. Other-ASD vs. Non-ASD diagnoses; SA-CSS: $F(2,2192) = 974.43, p < .001$; RRB-CSS: $F(2,2192) = 421.35, p < .001$). Nonetheless, there was marked overlap in the distribution of scores across the three diagnostic groups (see Fig. 3a, b).

Correlations Between Domain Calibrations and Overall Calibrated Severity Score

In the ASD sample, associations between SA-CSS and RRB-CSS were significant, but weak ($r = .25$; Cohen 1988). Although correlations between each of the domain calibrated scores and the overall CSS were both strong, the association between SA-CSS and CSS ($r = .89$) was greater than that observed for RRB-CSS and CSS ($r = .57$). This is a reflection that the overall total from which the CSS is derived is comprised of a greater proportion of items from the SA domain than the RRB domain.

Predictors of SA-Raw and SA-CSS

The final model including all predictors explained a total of 45 % of variance in the SA-Raw total. Verbal IQ and maternal education (mothers with graduate/professional degrees vs. all others) emerged as significant predictors of SA-Raw. In contrast, the same model accounted for only 13 % of the variance in the SA-CSS, with verbal IQ and nonverbal IQ both making small, but significant contributions to the calibrated SA score. Thus, although there is still a significant association between SA-CSS and the child's cognitive level, the calibrated SA scores are markedly less influenced by child cognitive level than SA-Raw.

Next, verbal IQ, nonverbal IQ, and maternal education were entered into a Forward Stepwise model to assess the relative contributions of each of these variables in predicting SA-Raw. As shown in Table 4, verbal IQ accounted for the majority of variance (43 %) and the contributions of nonverbal IQ and maternal education were minimal (0.3 and 0.2 %, respectively). In the Forward model predicting SA-CSS, verbal IQ accounted for 10.5 % of variance while nonverbal IQ explained an additional 0.4 %; maternal education was excluded by the model, indicating that it was not significant (see Table 4). These results reflect a reduction in the influence of verbal IQ from a large effect on SA-Raw ($R = .66$) to a small-to-medium effect on SA-CSS ($R = .33$; Cohen 1988; McCarthy et al. 1991). It is noteworthy that verbal and nonverbal IQ were highly correlated ($r = .76$) and when verbal IQ was removed as a predictor, nonverbal IQ accounted for 21.8 % of variance in SA-Raw and only 4.3 % in SA-CSS; both models excluded maternal education as a predictor.

Predictors of RRB-Raw and RRB-CSS

Child characteristics such as IQ explained much less variance in raw RRB totals (i.e., 15.3 %). Significant predictors included verbal IQ, nonverbal IQ, and race (African American vs. all others). In the Forward Stepwise Model, verbal IQ, nonverbal IQ and race each remained significant predictors of RRB-Raw (see Table 4). Verbal IQ accounted for the majority of variance (11.7 %) and nonverbal IQ and race each made small contributions (1.4 and 1.1 %, respectively). Again, if verbal IQ was excluded from the models, nonverbal IQ explained 11.4 % and race explained 0.8 % of variance in RRB-Raw.)

Calibrated RRB scores reduced the influence of child characteristics; in the end, child characteristics explained only 5.5 % of the variance, with verbal IQ, nonverbal IQ and race emerging as small, but significant predictors of RRB-CSS. In the Forward Model predicting RRB-CSS, nonverbal IQ explained 3.5 % of the variance in RRB-CSS; verbal IQ and race accounted for an additional 0.5 and 0.6 %, respectively.

Case Summaries

Four children with ASD diagnoses were chosen to demonstrate the utility of the newly calibrated domain scores for separately examining the severity of social and repetitive behaviors over time (see Table 5 for child characteristics at first and last assessments). Each child's SA-CSS and RRB-CSS are plotted by age in Fig. 4. Overall CSS scores are also provided; in many cases the overall CSS and SA-CSS follow similar, if not identical, trajectories, again reflecting that the overall total from which the CSS is derived is comprised of a greater proportion of items from the SA domain than the RRB domain.

Case 1—"Bianca," a Caucasian female, was diagnosed with autism at 4 years of age when she was first seen as a clinical referral (see Gotham et al. 2009). Her overall CSS suggests that her symptom severity was relatively stable across early childhood, followed by gradual a decrease in severity throughout late childhood and early adolescence. Her SA-CSS follows a similar trajectory, reflecting persistent difficulties with eye contact and unusual social overtures accompanied by an increase in use of gestures and shared enjoyment with the examiner. In contrast, her RRB-CSS follows a quite different pattern, with a RRB-CSS of 10 at Bianca's first assessment (reflecting her exhibition of sensory-seeking behaviors, delayed echolalia, repetitive asking of questions and repeated lining up of toys). This was followed by a considerable decrease in severity at age 5 and a year of relative stability, during which time she demonstrated some repetitive speech and mild preoccupations with a particular musician, but no hand and finger mannerisms. Although Bianca did not demonstrate repetitive behaviors when she was assessed at 8 years old, in early adolescence, she again exhibited clear hand and finger mannerisms and engaged in somewhat repetitive speech (though recall that there is not a RRB-CSS of 2–4, so the fluctuation in severity later childhood may appear greater than it actually was).

Case 2—"Joey," a Caucasian male, was first seen as a clinical referral at 2 years, 10 months of age, at which time he received a diagnosis of PDD-NOS. When first seen, he exhibited severe social-communication symptoms (i.e., an SA-CSS of 10 demonstrating poor eye contact and very limited social overtures), but mild repetitive behaviors (RRB-CSS of 5 reflecting very brief repetitive behaviors) during the ADOS. In his subsequent assessments, there was an apparent increase in repetitive behaviors due to his use of stereotyped language (e.g., "That's all folks!"), accompanied by an improvement in the social affect domain (i.e., improvements in eye contact and more frequent and appropriate overtures). At age 7 years, 7 months, Joey's SA-CSS of 3 and RRB-CSS of 7 suggested milder severity of social-communication symptoms compared to repetitive behaviors. His overall CSS followed a similar trajectory to his SA-CSS, showing a steady decrease in severity across early childhood, and did not reflect the apparent increase in repetitive behaviors during this same period.

Case 3—"Carolyn," a Caucasian female, was first seen as part of a clinical research project just after her second birthday. At this time, she received a diagnosis of PDD-NOS and her SA-CSS of 4 suggested milder severity of social-communication impairments during the ADOS (e.g., strengths in shared enjoyment and facial expressions, but difficulties using coordinated eye gaze) compared to her RRB-CSS of 9 (reflecting hand and finger, as well as

whole-body mannerisms, a preoccupation with cars and brief peering at objects). However, over the next 8 years, there was a steady increase in deficits in SA, resulting in an SA-CSS of 10 by the time she was 10 years old; while she continued to express some shared enjoyment with the examiner, her use of facial expressions was more limited and deficits in eye contact persisted. Her overall CSS also follows this pattern. In contrast, during the period in which she had the most dramatic increases in SA-CSS, the severity of Carolyn's repetitive behaviors remained relatively stable. Over time, she continued to exhibit hand and finger and whole-body mannerisms (e.g., twirling and jumping), and brief visual sensory interests. She also demonstrated unusual preoccupations (e.g., with time), as well as ritualistic behaviors, such as placing objects in toy trucks in a particular way.

Case 4—"Matthew," an African American male, was seen at age 4 years as part of a clinical research study, at which time he received a diagnosis of autism. During his first ADOS, Matthew exhibited more severe social-communication symptoms (SA-CSS = 8) than repetitive behaviors (RRB-CSS = 5). Separate examination of his SA-CSS and RRB-CSS suggest relatively stable severity in both domains across early childhood, marked by persistent difficulties in nonverbal social communication (e.g., facial expressions and eye contact), initiation of overtures, brief sensory interests and possible hand and finger mannerisms. At 11 years of age, Matthew showed an apparent decrease in severity of social-communication symptoms (a greater range of facial expressions and more reciprocal social communication) and a worsening of repetitive behaviors, including clear hand and finger mannerisms, excessive references to Batman and wrestling, repetitive stereotyped questions, and listing of his classmates when asked the names of his friends. In his case, the overall CSS showed a gradual worsening of symptom severity between ages 4 and 11, failing to account for the possible divergence of trajectories in social-communication skills and repetitive behaviors in later childhood.

Discussion

ADOS calibrated domain totals achieved the goal of significantly reducing associations with child characteristics compared to raw SA and RRB totals. For SA-Raw domain scores, 45 % of variance was explained by child characteristics not specific to ASD, with verbal IQ and maternal education emerging as significant predictors. For the SA-CSS, verbal IQ remained the only significant predictor, accounting for just under 11 % of variance in the calibrated SA score. Similarly, approximately 12 % of variance in RRB-Raw Total was explained by verbal IQ, with nonverbal IQ and race collectively accounting for an additional 3 %. For the RRB-CSS, nonverbal IQ, verbal IQ and race remained significant predictors, but explained less than 5 % of variance. Thus, though the effects of child characteristics were not completely eliminated, the calibrated domain scores provided a measure of ASD severity that was significantly less influenced by child characteristics, particularly verbal IQ, than were raw totals.

It is interesting to note that associations between IQ and RRB Raw were much smaller compared to the relationship between IQ and SA-Raw. A similar difference in associations with developmental level was noted for Social + Nonverbal Communication vs. Repetitive Behavior raw domain totals on the Autism Diagnostic Interview-Revised (Hus and Lord in

press). The restricted range of RRB-Raw scores may explain the weaker associations. Nevertheless, in spite of relatively smaller influences of developmental level on RRB-Raw, it is important to calibrate RRB-CSS in order to provide a comparable severity metric for both ADOS domains. Most important, the RRB-CSS reduced the influence of developmental level on RRB totals even further.

It is also noteworthy that there was marked overlap in the distributions of domain calibrated scores across diagnostic groups. On one hand, the overlap of the Non-ASD group with the Autism and Other-ASD groups may reflect recruitment biases in our Non-ASD sample, some of whom were referred for assessment of ASD, but who received a clinical Non-ASD diagnosis. On the other hand, the overlap between the Autism and Other-ASD group could reflect that the calibrated scores are capturing the heterogeneity of symptom severity that characterizes ASDs. Moreover, the overlap with the Non-ASD group highlights that some social-communication and repetitive behaviors captured on the ADOS are not specific to ASD.

The newly standardized SA-CSS and RRB-CSS provide useful measures of autism symptom severity which are consistent with the two symptom domains defining ASD proposed for DSM-5. As we move toward a single classification of “autism spectrum disorder” in DSM-5, calibrated domain scores have the potential to play a role in the clinical specification of ASD severity. When DSM-5 criteria are finalized, assessing the degree to which the 10-point CSS scale indicating severity of ASD symptomatology relates to different DSM-5 levels of severity for each behavioral domain (currently proposed as “requiring support,” “requiring substantial support,” and “requiring very substantial support”) will be an important step. If the scores can be mapped on to clinical levels of severity, they may be useful to inform the level of impairment in each behavioral domain; however, these scores will not be sufficient to make such clinical determinations, as they provide information about behaviors in a limited context. Information collected from other modalities of assessment, such as caregiver interview or observation in other settings, will be needed to inform the appropriate level of severity to describe the level of support an individual requires.

It is also hoped that the calibration of severity metrics for social-communication deficits (SA-CSS) and repetitive behaviors (RRB-CSS) will bring us a step closer to parsing apart the phenotypic heterogeneity in ASD. Current studies frequently rely on totals from diagnostic instruments such as the SRS or ADI-R as estimates of ASD severity. Yet these totals are known to be greatly influenced by child characteristics such as age, language level, and non-ASD-specific behavioral problems (e.g., Constantino et al. 2003; Hus et al. in press; Hus and Lord in press). Although the original ADOS calibrated severity metric was derived to reduce the effects of non-specific child characteristics (Gotham et al. 2009), it yields an estimate of overall severity that does not allow for separate examination of the variation in behavioral domains underlying these scores. In comparison, the SA-CSS and RRB-CSS provide more behavioral specificity than each of these general measures. Because potential biomarkers are frequently postulated to be related to specific domains of behavior (e.g., severity of RRBs), separate calibrated domain scores offer an important advance. Additionally, use of these calibrated domain scores in place of raw totals increases the

likelihood that associations with genetic or neurobiological abnormalities are specific to ASD symptoms rather than associated with general developmental factors, such as age, IQ or language level.

Using these scores to separately examine distinct trajectories of social-communication and repetitive behaviors may also provide a more sensitive measure of intervention response over longer periods of time, enabling change in one domain to be detected, even when behaviors in the second domain persist. Although children may become more familiar with particular tasks (e.g., participating in the birthday party routine) if they are administered the ADOS several times within a short period, because scores are based on spontaneous initiations and responses, rather than performance on tasks, scores and ADOS classifications do not demonstrate practice effects (Lord et al. 2012). Thus, the SA-CSS and RRB-CSS may provide a way to measure more global changes in behaviors in response to intervention, rather than improvements in very specific skills. Furthermore, different SA-CSS and RRB-CSS trajectory profiles may provide an additional method of stratification to increase phenotypic homogeneity in samples, which can be used to gain insight into biological mechanisms underlying specific developmental patterns.

Limitations

Domain calibrations were based on the large “convenience” sample that was used to create the overall ADOS CSS (Gotham et al. 2009). As these authors acknowledged, this sample is likely to be representative of other samples ascertained through North American clinical research centers over the past two decades. It is hoped that using ADOS classifications of (i.e., “Autism,” “Autism Spectrum” and “NonSpectrum”), rather than clinical best estimate diagnoses, to anchor overall severity scores and set thresholds for sensitivity and specificity of domain calibrated scores would circumvent, to some extent, recruitment effects in this sample (Gotham et al. 2009). However, it is possible that calibration using ADOSes from population studies or more recently ascertained samples may result in different mappings of raw totals and calibrated scores. Additionally, samples recruited outside of North America, or from other clinical populations, may show a somewhat different distribution of scores. Here, the effects of maternal education and race observed on both overall raw totals and calibrated scores are likely to be an artifact of recruitment biases (Gotham et al. 2009), though the significance of these predictors may also have been influenced by the large sample size. Replication of the domain calibrations in independent samples is an important next step.

Given the restricted range of raw RRB totals, the RRB-CSS is not a full 10-point severity metric. Nonetheless, scores were mapped onto the 10-point scale to avoid confusion when using the calibrated domain scores together. That is, there was concern that a reduced RRB-CSS scale (e.g., of 1–6) may result in confusion when interpreting the meaning of an RRB-CSS score in comparison to a score on the overall-CSS scale (i.e., an assumption that an RRB-CSS of 6 would be equal to an overall-CSS of 6, when it actually would be more meaningful to interpret as similar to an overall-CSS of 10). The method of using the overall-CSS percentiles to inform mapping of domain raw scores to the 10-point calibrated scale allows comparability across the three scales, such that a given value on the overall-CSS, SA-

CSS, and RRB-CSS correspond to approximately the same percentile of raw score (for a child of that language level and age) for each. Such comparability also increases the clinical utility of this metric; for example, a child who has a high overall-CSS comprised of an SA-CSS of '10' and an RRB-CSS of '6' may need a different treatment approach than another child with the same overall-CSS reflecting an RRB-CSS of '10' and an SA-CSS of '6'. When using scores to monitor change over time or in response to intervention, researchers and clinicians must bear in mind that there are not RRB-CSS values of 2, 3 or 4. Thus, changing from a score from RRB-CSS of 1, indicating that no repetitive behaviors were observed during the ADOS, to 5 (reflecting mild severity), is not the same as a change in severity from an RRB score of 6–10. This distribution of scores reflects that, given the limited timeframe of the ADOS, the presence of repetitive behaviors is likely to be more meaningful than the absence of such. In order to ensure that a change in CSS for either domain is meaningful, the lower (or higher) score should be observed across several time points. In contrast, a significant increase or decrease during one particular session may suggest that other factors were influencing the child's behavior on that particular day.

Conclusions

ADOS domain calibrations provide separate estimates of severity of ASD-related social-communication deficits and repetitive behaviors that are relatively independent of child characteristics, such as age and language skills, compared to their respective raw totals. This improves their utility as continuous measures of ASD symptom severity that can be used to increase homogeneity of samples and identify links between specific behavioral domains and biological mechanisms, as well as to examine different trajectories of ASD symptoms over time.

Acknowledgments

This research was supported by a Dennis Weatherstone Predoctoral Fellowship to VH and National Institute of Mental Health grants T32-MH18921 to KG and R01MH081873 and RC1MH089721 to CL. We gratefully acknowledge Drs. Andrew Pickles, Christopher Gruber and Sheri Stegall for their consultation in preparation of this manuscript, as well as all of the families who participated in this research.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Fourth Edition. American Psychiatric Association; Washington, DC: 2000. Text Revision: DSM-IV-TR
- American Psychiatric Association. [Accessed 04 March 2012] DSM-5 Development: A 09 Autism Spectrum Disorder. 2011. <http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=94>
- Cohen, J. Statistical power analysis for the behavioral sciences. Psychology Press; 1988.
- Constantino, JN.; Gruber, C. The Social Responsiveness Scale. Western Psychological Services; Los Angeles: 2005.
- Constantino JN, Hudziak JJ, Todd RD. Deficits in reciprocal social behavior in male twins: evidence for a genetically independent domain of psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2003; 42(4):458–467. doi:10.1097/01.CHI.0000046811.95464.21. [PubMed: 12649633]
- Di Martino A, Kelly C, Grzadzinski R, Zuo X-N, Mennes M, Mairena MA, et al. Aberrant striatal functional connectivity in children with autism. *Biological Psychiatry*. 2011; 69(9):847–856. doi: 10.1016/j.biopsych.2010.10.029. [PubMed: 21195388]

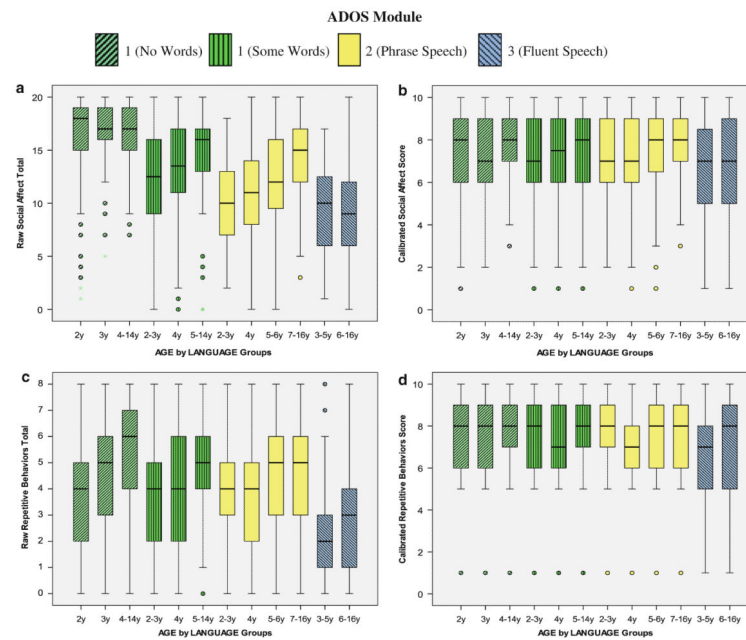
- Dichter GS, Richey JA, Rittenberg AM, Sabatino A, Bodfish JW. Reward circuitry function in autism during face anticipation and outcomes. *Journal of Autism and Developmental Disorders*. 2011 doi: 10.1007/s10803-011-1221-1.
- DiLavore P, Lord C, Rutter M. The Pre-Linguistic Autism Diagnostic Observation Schedule. *Journal of Autism and Developmental Disorders*. 1995; 25(4):355–379. doi:10.1007/BF02179373. [PubMed: 7592249]
- Elliott, CD. *Differential Ability Scales*. Psychological Corporation; San Antonio: 1990.
- Frazier TW, Youngstrom EA, Speer L, Embacher R, Law P, Constantino J, et al. Validation of proposed DSM-5 criteria for Autism Spectrum Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2012; 51(1):28–40. doi: 10.1016/j.jaac.2011.09.021. [PubMed: 22176937]
- Gotham K, Risi S, Pickles A, Lord C. The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*. 2007; 37(4):613–627. doi:10.1007/s10803-006-0280-1. [PubMed: 17180459]
- Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2009; 39(5):693–705. doi: 10.1007/s10803-008-0674-3. [PubMed: 19082876]
- Huerta M, Bishop SL, Duncan A, Hus V, Lord C. How many will we miss? Application of DSM 5 criteria to 3 samples of children with DSM-IV diagnoses of ASD. *American Journal of Psychiatry*. (in press).
- Hus V, Lord C. Effects of child characteristics on the Autism Diagnostic Interview-Revised: Implications for use of scores as a measure of ASD severity. *Journal of Autism and Developmental Disorders*. (in press). doi:10.1007/s10803-012-1576-y.
- Hus V, Bishop S, Gotham K, Huerta M, Pickles A, Lord C. Factors influencing scores on the Social Responsiveness Scale. *Journal of Child Psychology and Psychiatry*. (in press).
- Lord C, Petkova E, Hus V, Gan W, Lu F, Martin DM, et al. A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*. 2011 doi:10.1001/archgenpsychiatry.2011.148.
- Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A. Autism from 2 to 9 years of age. *Archives of General Psychiatry*. 2006; 63:694–701. doi:10.1001/archpsyc.63.6.694. [PubMed: 16754843]
- Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S.; Gotham, K.; Bishop, S. *Autism Diagnostic Observation Schedule*. 2nd Edition. Western Psychological Services; Torrance: 2012. (ADOS-2)
- Luyster R, Gotham K, Guthrie W, Coffing M, Petrak R, Pierce K, et al. The autism diagnostic observation schedule—toddler module: a new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2009; 39(9):1305–1320. doi:10.1007/s10803-009-0746-z. [PubMed: 19415479]
- Mandy WPL, Charman T, Skuse DH. Testing the construct validity of proposed criteria for DSM-5 Autism Spectrum Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2012; 51(1):41–50. doi:10.1016/j.jaac.2011.10.013. [PubMed: 22176938]
- Mattila M-L, Kielinen M, Linna S-L, Jussila K, Ebeling H, Bloigu R, Joseph RM, et al. Autism Spectrum Disorders according to “DSM-IV-TR” and comparison with “DSM-5” draft criteria: An epidemiological study. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2011; 50(6):583–592.e11. [PubMed: 21621142]
- McCarthy PL, Cicchetti DV, Sznajderman SD, Forsyth BC, Baron MA, Fink HD, et al. Demographic, clinical, and psychosocial predictors of the reliability of mothers’ clinical judgments. *Pediatrics*. 1991; 88(5):1041–1046. [PubMed: 1945609]
- McPartland JC, Reichow B, Volkmar FR. Sensitivity and specificity of proposed DSM-5 diagnostic criteria for Autism Spectrum Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2012; 51(4):368–383. doi:10.1016/j.jaac.2012.01.007. [PubMed: 22449643]
- Mullen, E. *The Mullen Scales of Early Learning*. American Guidance Service, Inc; Circle Pines: 1995.
- Rutter, M.; Le Couteur, A.; Lord, C. *Autism diagnostic interview-revised*. Western Psychological Services; Los Angeles: 2003.

- State MW, Levitt P. The conundrums of understanding genetic risks for autism spectrum disorders. *Nature Neuroscience*. 2011; 14(12):1499–1506. doi:10.1038/nn.2924.
- World Health Organization. The ICD-10 classification of mental and behavioral disorder: Clinical descriptions and diagnostic guidelines. World Health Organization; Geneva: 1992.
- World Health Organization. [Accessed 22 September 2012] International Classification of Diseases 11th Revision: Beta draft. 2012. <http://apps.who.int/classifications/icd11/browse/f/en>

| Age (in years) | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10+ |
|----------------|--------------|-------|-------|-------|-------|-------|-------|---|---|-----|
| Module 1 | No Words | n=203 | n=141 | n=216 | | | | | | |
| | Single Words | n=214 | | n=82 | n=108 | | | | | |
| Module 2 | Phrases | n=106 | | n=94 | n=103 | | n=112 | | | |
| Module 3 | Fluent | n=71 | | | | n=357 | | | | |

Fig. 1.

Age by language level calibration cells. Note. *Ns* denote the number of ASD assessments within each cell

**Fig. 2.**

a (*top, left*) Distributions of raw Social Affect domain totals by age/language cells. **b** (*top, right*) Distributions of calibrated Social Affect domain scores by age/language cells. **c** (*bottom, left*) Distributions of raw Restricted and Repetitive Behavior domain totals by age/language cells. **d** (*bottom, right*) Distributions of calibrated Restricted and Repetitive Behavior domain scores by age/language cells

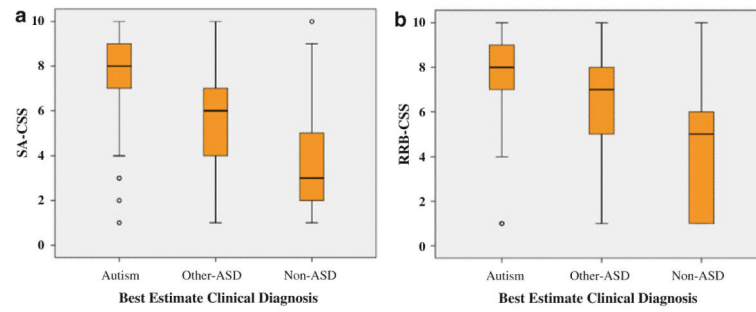


Fig. 3.

a (*left*) Distributions of calibrated Social Affect domain scores by best estimate clinical diagnosis. **b** (*right*) Distributions of calibrated Restricted and Repetitive Behavior domain scores by best estimate clinical diagnosis

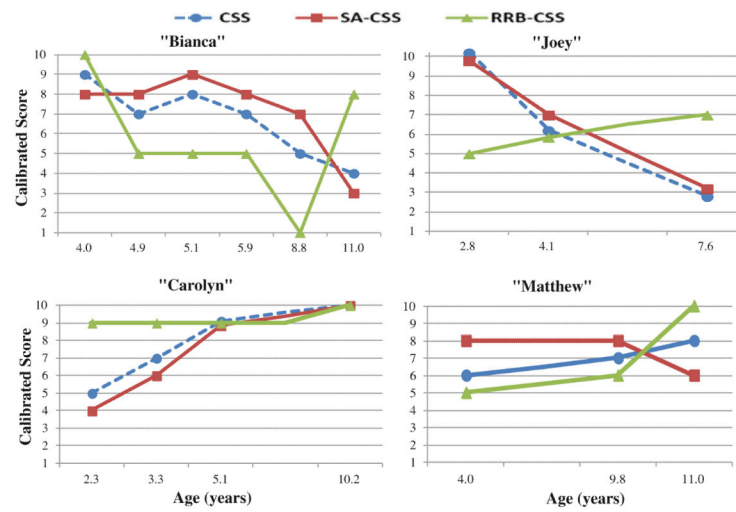


Fig. 4.
Case summaries of longitudinal domain severity scores

Table 1

Sample descriptives

| | Module 1, no words | | | Module 1, some words | | | Module 2, younger than 5 | | | Module 2, 5 or older | | | Module 3 | | |
|----------------|--------------------|-------|-------|----------------------|-------|-------|--------------------------|-------|-------|----------------------|-------|-------|----------|-------|-------|
| | N | Mean | SD | N | Mean | SD | N | Mean | SD | N | Mean | SD | N | Mean | SD |
| ASD | | | | | | | | | | | | | | | |
| Age | 551 | 4.22 | 2.21 | 395 | 4.41 | 1.99 | 197 | 3.78 | .78 | 215 | 7.82 | 2.54 | 428 | 8.54 | 2.54 |
| VIQ | 522 | 26.84 | 14.71 | 361 | 52.63 | 21.80 | 164 | 80.80 | 20.93 | 199 | 55.14 | 19.49 | 386 | 95.53 | 22.97 |
| NVIQ | 515 | 53.16 | 21.40 | 358 | 69.74 | 21.67 | 161 | 92.53 | 22.82 | 201 | 76.60 | 23.43 | 383 | 96.22 | 22.32 |
| VMA | 528 | .97 | 1.49 | 355 | 2.32 | 3.01 | 163 | 4.45 | 6.05 | 202 | 4.63 | 4.60 | 377 | 8.46 | 5.36 |
| NVMA | 516 | 1.98 | .83 | 359 | 3.05 | 2.40 | 158 | 3.69 | 1.35 | 190 | 5.74 | 2.25 | 357 | 8.23 | 2.88 |
| SA Raw | 551 | 16.79 | 2.95 | 395 | 13.23 | 4.44 | 197 | 10.44 | 4.30 | 215 | 13.20 | 4.29 | 428 | 9.26 | 4.37 |
| RRB Raw | 551 | 4.67 | 2.11 | 395 | 4.07 | 2.07 | 197 | 3.90 | 2.11 | 215 | 4.68 | 2.10 | 428 | 2.71 | 1.87 |
| Non-ASD | | | | | | | | | | | | | | | |
| Age | 60 | 3.30 | 1.61 | 107 | 3.51 | 1.60 | 57 | 3.67 | .62 | 44 | 8.00 | 2.55 | 141 | 8.95 | 2.47 |
| VIQ | 57 | 40.96 | 18.72 | 90 | 68.08 | 23.74 | 51 | 85.33 | 21.83 | 44 | 58.09 | 19.06 | 135 | 91.70 | 22.29 |
| NVIQ | 55 | 58.80 | 28.73 | 89 | 70.52 | 23.75 | 49 | 92.04 | 20.46 | 44 | 61.93 | 24.13 | 136 | 89.85 | 22.23 |
| VMA | 57 | 1.15 | .47 | 87 | 2.30 | .69 | 50 | 4.72 | 6.29 | 43 | 4.18 | 1.16 | 134 | 8.60 | 5.13 |
| NVMA | 55 | 1.72 | .72 | 86 | 2.44 | .71 | 46 | 3.47 | .83 | 44 | 4.72 | 1.39 | 132 | 7.92 | 2.71 |
| SA Raw | 60 | 8.37 | 5.83 | 107 | 4.71 | 3.91 | 57 | 3.56 | 2.77 | 44 | 4.16 | 3.14 | 141 | 3.90 | 2.95 |
| RRB Raw | 60 | 1.88 | 1.88 | 107 | 1.40 | 1.49 | 57 | 1.49 | 1.43 | 44 | 1.64 | 1.64 | 141 | .99 | 1.15 |

ASD autism spectrum disorder (Autistic Disorder, Aspergers, PDD-NOS); VIQ verbal IQ; NVIQ nonverbal IQ; VMA nonverbal mental age; NVMA nonverbal mental age; SA Raw Social Affect raw total, RRB Raw Restricted, Repetitive Behaviors raw total, Non-ASD non autism spectrum disorder diagnosis

Table 2

Mapping of ADOS raw domain totals onto calibrated severity scores

| Domain | Calibrated severity score | Raw domain totals | | | | | | | | | | | |
|--|---------------------------|--------------------|---------|------------|----------------------|---------|------------|-----------|---------|-----------|------------|-----------|------------|
| | | Module 1; no words | | | Module 1; some words | | | Module 2 | | | | Module 3 | |
| | | 2 years | 3 years | 4–14 years | 2–3 years | 4 years | 5–14 years | 2–3 years | 4 years | 5–6 years | 7–16 years | 3–5 years | 6–16 years |
| Social affect domain | 1 | 0–3 | 0–3 | 0–2 | 0–1 | 0–1 | 0–1 | 0–1 | 0–1 | 0–1 | 0–1 | 0–2 | 0–1 |
| | 2 | 4–5 | 4–5 | 3–5 | 2–4 | 2–3 | 2–3 | 2–3 | 2 | 2–3 | 2 | 3 | 2 |
| | 3 | 6–8 | 6–9 | 6–9 | 5 | 4–5 | 4–5 | 4 | 3–4 | 4–5 | 3–4 | 4 | 3–4 |
| | 4 | 9 | 10 | 10 | 6–7 | 6–7 | 6–7 | 5 | 5–6 | 6 | 5 | 5 | 5 |
| | 5 | 10–13 | 11–12 | 11–12 | 8 | 8–9 | 8–9 | 6 | 7 | 7 | 6–7 | 6 | 6 |
| | 6 | 14–16 | 13–16 | 13–14 | 9–11 | 10–12 | 10–13 | 7–8 | 8–9 | 8–9 | 8–10 | 7–8 | 7 |
| | 7 | 17 | 17 | 15–16 | 12–13 | 13 | 14–15 | 9–10 | 10–11 | 10–11 | 11–13 | 9–10 | 8–9 |
| | 8 | 18 | 18 | 17–18 | 14–15 | 14–15 | 16 | 11 | 12–13 | 12–15 | 14–15 | 11–12 | 10–11 |
| | 9 | 19 | 19 | 19 | 16–17 | 16–17 | 17–18 | 12–14 | 14–15 | 16 | 16–17 | 13–14 | 12–14 |
| | 10 | 20 | 20 | 20 | 18–20 | 18–20 | 19–20 | 15–20 | 16–20 | 17–20 | 18–20 | 15–20 | 15–20 |
| Restricted and Repetitive behaviors domain | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 2 | | | | | | | | | | | | |
| | 3 | | | | | | | | | | | | |
| | 4 | 1 | 1 | 1–2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 5 | | | | | | | | | | | | |
| | 6 | 2 | 2–3 | 3 | 2 | 2 | 2–3 | 2 | 2–3 | 2–3 | 2–3 | | |
| | 7 | 3 | 4 | 4 | 3 | 3–4 | 4 | 3 | 4 | 4 | 4 | 2 | 2 |
| | 8 | 4 | 5 | 5–6 | 4 | 5 | 5 | 4 | 5 | 5 | 5 | 3 | 3 |
| | 9 | 5 | 6 | 7 | 5 | 6 | 6 | 5–6 | 6 | 6 | 6 | 4 | 4–5 |
| | 10 | 6–8 | 7–8 | 8 | 6–8 | 7–8 | 7–8 | 7–8 | 7–8 | 7–8 | 7–8 | 5–8 | 6–8 |

Table 3

Domain raw totals and calibrated severity score means and standard deviations by age/language cell (ASD assessments only)

| Module | Age (years) | N | SA-Raw | | SA-CSS | | RRB-Raw | | RRB-CSS | |
|-----------------------|-------------|------|--------|------|--------|------|---------|------|---------|------|
| | | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Module 1, no words | 2 | 203 | 16.38 | 3.85 | 7.36 | 2.11 | 3.75 | 2.01 | 7.49 | 2.21 |
| | 3 | 141 | 16.88 | 2.88 | 7.45 | 1.75 | 4.76 | 2.00 | 7.77 | 1.75 |
| | 4-14 | 216 | 16.76 | 2.59 | 7.75 | 1.46 | 5.36 | 2.03 | 7.82 | 1.65 |
| Module 1, some words | 2-3 | 214 | 12.10 | 4.73 | 6.81 | 2.33 | 3.66 | 1.94 | 7.44 | 2.10 |
| | 4 | 82 | 13.01 | 4.75 | 7.16 | 2.39 | 4.12 | 2.34 | 7.30 | 2.25 |
| | 5-14 | 108 | 14.85 | 3.65 | 7.57 | 1.78 | 4.67 | 2.01 | 7.56 | 2.05 |
| Module 2 | 2-3 | 106 | 10.03 | 4.02 | 7.08 | 2.18 | 4.02 | 2.02 | 7.59 | 1.94 |
| | 4 | 94 | 10.69 | 4.69 | 6.88 | 2.37 | 3.74 | 2.18 | 6.87 | 2.22 |
| | 5-6 | 103 | 12.25 | 4.62 | 7.49 | 2.05 | 4.59 | 2.09 | 7.59 | 1.93 |
| Module 3 | 7-16 | 112 | 14.07 | 3.79 | 7.99 | 1.59 | 4.77 | 2.11 | 7.67 | 2.02 |
| | 3-5 | 71 | 9.52 | 4.06 | 6.68 | 2.48 | 2.65 | 1.83 | 6.94 | 2.43 |
| | 6-16 | 357 | 9.21 | 4.43 | 6.77 | 2.52 | 2.73 | 1.88 | 6.86 | 2.68 |
| All modules, all ages | | 1807 | 12.98 | 4.99 | 7.21 | 2.17 | 3.96 | 2.18 | 7.39 | 2.19 |

Table 4

Forward stepwise linear regression models for domain raw totals and calibrated domain scores

| | SA-Raw | | | | | | SA-CSS | | | | | |
|--------------|----------------|---------|------|-------|------|------|----------------|--------|------|------|------|------|
| | R ² | F | df | B | SE B | β | R ² | F | df | B | SE B | β |
| Step 1 | .430 | 1079.07 | 1430 | | | | .105 | 167.72 | 1430 | | | |
| Constant | | | | 18.75 | .20 | | | | | 8.45 | .11 | |
| Verbal IQ | | | | -.10 | .00 | -.66 | | | | -.02 | .00 | -.32 |
| Step 2 | .433 | 7.41 | 1429 | | | | .109 | 5.97 | 1429 | | | |
| Constant | | | | 18.19 | .29 | | | | | 8.18 | .16 | |
| Verbal IQ | | | | -.11 | .00 | -.72 | | | | -.03 | .00 | -.40 |
| Nonverbal IQ | | | | .02 | .01 | .08 | | | | .01 | .00 | .09 |
| Step 3 | .435 | 5.51 | 1428 | | | | | | | | | |
| Constant | | | | 18.15 | .29 | | | | | | | |
| Verbal IQ | | | | -.11 | .00 | -.73 | | | | | | |
| Nonverbal IQ | | | | .01 | .01 | .08 | | | | | | |
| Mat Ed | | | | .56 | .24 | .05 | | | | | | |

| | RRB-Raw | | | | | | RRB-CSS | | | | | |
|--------------|----------------|--------|------|------|------|------|----------------|-------|------|------|------|------|
| | R ² | F | df | B | SE B | β | R ² | F | df | B | SE B | β |
| Step 1 | .117 | 208.86 | 1573 | | | | .035 | 56.62 | 1573 | | | |
| Constant | | | | 5.30 | .10 | | | | | 8.49 | .15 | |
| Verbal IQ | | | | -.02 | .00 | -.34 | | | | -.02 | .00 | -.19 |
| Step 2 | .131 | 25.64 | 1572 | | | | .041 | 10.12 | 1572 | | | |
| Constant | | | | 5.83 | .15 | | | | | 8.68 | .16 | |
| Verbal IQ | | | | -.01 | .00 | -.20 | | | | -.02 | .00 | -.21 |
| Nonverbal IQ | | | | -.01 | .00 | -.18 | | | | -.50 | .16 | -.08 |
| Step 3 | .143 | 20.40 | 1571 | | | | .045 | 7.47 | 1571 | | | |
| Constant | | | | 6.07 | .15 | | | | | 8.64 | .16 | |
| Verbal IQ | | | | -.01 | .00 | -.23 | | | | -.01 | .00 | -.13 |
| Nonverbal IQ | | | | -.02 | .00 | -.19 | | | | -.56 | .16 | -.09 |
| Race | | | | -.67 | .15 | -.11 | | | | -.01 | .00 | -.11 |

Table 5

Case summary characteristics

| Demographics | | | First assessment | | | | Last assessment | | | | |
|---------------------|--------|-----------|------------------|-----|------|-------------|-----------------|------|------|-------------|---|
| Gender | Race | Diagnosis | Age | VIQ | NVIQ | ADOS module | Age | VIQ | NVIQ | ADOS module | |
| Bianca ^d | Female | White | Autism | 4.0 | 108 | 80 | 2 | 11.0 | 126 | 107 | 3 |
| Joey | Male | White | PDD-NOS | 2.8 | 69 | 74 | 2 | 5.1 | 105 | 119 | 3 |
| Carolyn | Female | White | PDD-NOS | 2.3 | 33 | 72 | 1 | 10.2 | 42 | 51 | 2 |
| Matthew | Male | Af. Amer. | Autism | 4.0 | 31 | 63 | 1 | 11.0 | 58 | 88 | 3 |

All ages in years; VIQ verbal IQ, NVIQ nonverbal IQ

^dCognitive assessment was not completed at last assessment; IQs are from previous assessment at age 10