



Published in final edited form as:

Autism. 2019 July ; 23(5): 1176–1185. doi:10.1177/1362361318793253.

Measuring Treatment Response in Children with Autism Spectrum Disorder (ASD): Applications of the Brief Observation of Social Communication Change (BOSCC) to the Autism Diagnostic Observation Schedule (ADOS)

So Hyun Kim, Ph.D.¹, Rebecca Grzadzinski, Ph.D.², Kassandra Martinez, B.A.¹, and Catherine Lord, Ph.D.¹

¹Department of Psychiatry, Weill Cornell Medicine, White Plains, NY, 10601

²Teachers College, Columbia University, New York, NY, 10027

Abstract

This study aims to determine the validity and reliability of applying the coding strategy from the Brief Observation of Social Communication Change (BOSCC), a newly validated treatment outcome measure, to videotaped segments of the Autism Diagnostic Observation Schedule (ADOS). Results indicate strong reliability and validity of the BOSCC ratings using the ADOS segments in detecting changes in social communication over the course of treatment in young, minimally verbal children with ASD. Results also suggest that the BOSCC, when applied to ADOS segments, may be more sensitive in detecting subtle changes in social communication compared to the ADOS Calibrated Severity Scores (CSS). These results may support the application of the BOSCC to pre-existing datasets of ADOS videos to examine treatment responses.

Keywords

Autism Diagnostic Observation Schedule; Brief Observation of Social Communication Change; autism spectrum disorder; outcome measure

Many studies of early interventions focused on improving social communication in children with ASD have shown moderate changes in cognitive and language skills, but only minimal or no changes in core ASD symptoms (Green et al., 2010; Kasari et al., 2014; Rogers et al., 2012; Wetherby et al., 2014). This may be due to the lack of a treatment outcome measure that is sensitive enough to capture subtle changes in social communication symptoms over time. Due to the lack of uniform measurement approaches across different studies (Bolte & Diehl, 2013), replications and comparisons of results from various Randomized Control Trials (RCTs) have not been feasible. Evaluating response to treatment in ASD requires valid outcome measures that are sensitive enough to detect changes in the core symptoms of ASD that can be used across different RCTs. Previously, use of tools like the Autism

Diagnostic Observation Schedule (ADOS-2; Lord et al., 2012) were encouraged to measure behavioral changes over the course of treatment (Cunningham, 2012; Matson, 2007). However, researchers and clinicians have recently urged the field to move beyond using measures that were not designed to be treatment outcome measures, such as the ADOS, to evaluate treatment effects (Anagnostou et al., 2015).

In response to the need for a measure of treatment response that adequately captures changes in social communication, the Brief Observation of Social Communication Change (BOSCC; Grzadzinski et al., 2016) was recently developed and validated with a group of 56 minimally verbal children with ASD, ages 1–5 years. The BOSCC coding scheme was developed by expanding the codes of the ADOS to range from 0 to 5 in order to capture more nuanced behavioral changes that diagnostic codes may not adequately distinguish. This initial study demonstrated high inter-rater and test-retest reliability as well as convergent validity with measures of language and adaptive communication skills (Grzadzinski et al., 2016). In addition, in this initial study, the BOSCC Core total demonstrated statistically significant amounts of change over time compared to a no change alternative while the ADOS Calibrated Severity Scores (CSS; Gotham, Pickles, & Lord, 2009) over the same period of time did not. The BOSCC has shown promising evidence as a primary outcome measure for treatment response in a few other studies as well (Kitzerow, Teufel, Wilker, & Freitag, 2016; Pijl et al., 2016), though has not always yielded positive results (Fletcher-Watson et al., 2015).

The initial BOSCC psychometrics were drawn from videotaped parent-child play interactions (Grzadzinski et al., 2016). As the BOSCC has shown adequate validity and reliability based on parent-child interactions, it is important to explore whether the BOSCC coding scheme can be applied to other contexts, such as videotaped ADOS administrations. This may be especially useful for evaluating treatment efficacy using retrospective data from completed RCTs or other studies with existing videotaped ADOS administrations. While previous investigations may have shown minimal changes in ADOS CSS, the application of a more sensitive coding scheme, such as the BOSCC, to videotaped ADOS sessions may reveal additional evidence of behavioral changes in young, minimally-verbal children with ASD.

The goal of the current investigation is to provide evidence for the validity of the application of the BOSCC (referred to as “Standard BOSCC” hereafter) codes to videotaped ADOS segments (referred to as “ADOS-BOSCC” hereafter) for minimally verbal children. Specifically, by applying the Standard BOSCC coding scheme to segments from videotaped ADOS administrations, we aim to 1) determine if ADOS-BOSCC items capture variability in behaviors; 2) confirm the factor structure of the ADOS-BOSCC; 3) examine inter-rater and test-retest reliability of the ADOS-BOSCC; and 4) provide validity data for the ADOS-BOSCC in capturing changes in social communication over the course of treatment.

Method

Participants

Participants were drawn from clinically referred children for ASD who were invited to participate in early intervention through various RCTs (Kasari et al., 2014; Rogers et al., 2012; Wetherby et al., 2014). Of all the children who participated in these studies, we selected 49 children whose parent interaction videos and ADOS sessions were available within a 1- to 2-week time period to apply the Standard BOSCC and ADOS-BOSCC coding schemes respectively. Of the 49 children included in this study 10 (20%) were from Kasari et al. (2014), and 39 (80%) from Wetherby et al. (2014). Three (6%) of the children from the sample in Wetherby et al. (2014) also received the treatment outlined in Rogers et al. (2012). All of these data were collected at the University of Michigan Autism and Communication Disorders Center (UMACC). All children had best estimate clinical diagnoses of ASD based on diagnostic evaluations using the ADI-R (Lord, Rutter, & Couteur, 1994) and the ADOS2 (Lord et al., 2012a; Lord et al., 2012b) as well as developmental testing (Mullen, 1995). Because this work focuses on the validity and reliability of the ADOS-BOSCC, we did not explore effects of specific treatment conditions. The children in the study were between one and five years old at entry ($M=25.0$, $SD=9.7$) and a majority of children had limited spontaneous language (simple phrase speech or less; $n=37$ for ADOS Toddler Module, $n=9$ for ADOS Module 1, $n=3$ for ADOS Module 2 at Time 1). These children were followed for about 9 months on average ($M=8.8$, $SD=4.8$) with their final visit at the mean age of 3 years ($M=33.6$ months, $SD=9.8$ months). A majority of the children were still minimally verbal at Time 2 ($n=17$ for ADOS Toddler Module, $n=20$ for ADOS Module 1, $n=12$ for ADOS Module 2 at Time 2). See Table 1 for demographic and baseline characteristics.

Primary Measures

ASD Symptoms—Research reliable clinicians administered and scored the Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2; Lord et al., 2012a; Lord et al., 2012b) to all children for all time-points. The ADOS-2 yields overall Calibrated Severity Scores (CSS Total), Social Affect scores (CSS SA), and Restricted and Repetitive Behavior Scores (CSS RRB) (Gotham, Pickles, & Lord, 2009; Hus, Gotham, & Lord, 2014). The CSS has been found to be less affected by developmental factors such as IQ, language level, and age and allows for the comparison of scores across different ADOS modules compared to the ADOS total scores (Gotham et al., 2009; Hus, Gotham, & Lord, 2014).

BOSCC

Standard BOSCC.: The Standard BOSCC can be applied to 10 to 12-minute videos of parent- or examiner-child interaction videos. For the present study, ten-minute video observations of parent-child play interactions in the clinic were available from previously conducted intervention studies. Parents had been instructed to play with their child as they typically would at home. Consistent with the previously published BOSCC study (Grzadzinski et al., 2016), the Standard BOSCC coding scheme was applied to these interactions (Figure 1 & 2). The Standard BOSCC consists of 15 items that are coded on a 6-point scale ranging from 0 (the abnormality is not present) to 5 (the abnormality is present and it significantly impairs functioning). Thus, higher scores indicate more abnormality or

impairment. Items 1–8 focus on Social Communication (SC), while items 9–12 capture Restricted and Repetitive Behaviors (RRBs). The Core total combines the SC and RRB scores. Items 13–15 measure other abnormal behaviors often observed in individuals with ASD. A subset of the sample in the current study ($n=32$) overlapped with the sample of children who were included in the previous study examining psychometrics of the Standard BOSCC (Grzadzinski et al., 2016).

ADOS-BOSCC.: In addition to the parent-child play interaction, a videotaped ADOS was administered by a trained clinician. In the current study, the Standard BOSCC (with an additional Requesting code which was not included in the original coding scheme Grzadzinski et al., 2016) described above was applied to 12 minutes of videotaped ADOS administrations. In addition to the 15 items, one item was added to capture requesting since several ADOS probes provide standardized opportunities for the child to display requesting behaviors (Figure 1). The segments of the ADOS that the Standard BOSCC coding scheme was applied to were selected in a standardized way (Figure 2). The first segment included three minutes of Free Play and three minutes of Bubble Play. If the Free Play or Bubble Play segments were less than three minutes long (such that the total segment was less than 6 minutes), the remaining time was supplemented with part of the Response to Joint Attention (RJA) activity. RJA was added to 30% of the coded observations in this sample. Response to Name or Blocking Toy Play were excluded from the segment if they occurred during Free Play or Bubble Play. The second segment included three minutes of Birthday Party or Bath Time (depending on which ADOS module was administered) and three minutes of Anticipation of Routine with Objects. If either of these clips were less than three minutes long, Snack was added to the segment to have a total of 6 minutes. Snack was added to 56% of the coded observations in this sample. Ignoring during Bath Time was excluded from the segment. The selected clips were coded from the beginning of the task up to the designated time point, with the exception of Free Play, which was coded after the child was able to explore the toys for one minute. Coders were able to quickly identify the needed ADOS activities due to their familiarity with the ADOS tasks. On average, it took coders less than three minutes to identify the specific ADOS activities needed.

Because not all of the assessments occurred on the same day, each parent-child play interaction was matched with an ADOS observation that occurred within one week ($M=1.3$ days, $SD=5.7$ days). Between 2 and 5 matched pairs (Standard BOSCC and ADOS-BOSCC coding) were available per child with an average of 8.8 months ($SD=4.8$) between the first observation and last observation. At entry, children were between 16 and 54 months of age ($M=25.0$, $SD=9.7$) and between 18 and 61 months at exit ($M=33.6$, $SD=9.8$). Videos were coded by research assistants who had achieved an 80% inter-rater agreement standard on three consecutive training videos. All coders were blind to treatment type and time-point.

Additional Measures

Several assessments were gathered throughout the course of the intervention studies, such as assessments of adaptive and cognitive functioning. These measures were used in the current study to assess convergent validity of the ADOS-BOSCC.

Adaptive Functioning—To assess adaptive functioning, caregivers completed the Vineland Adaptive Behavior Scales (VABS; Sparrow, Cicchetti, & Balla, 2005). The VABS was available for 32 children at two time-points. The VABS yields standard scores in four domains: Communication, Socialization, Daily Living Skills, and Motor Skills. The rest of the children received the VABS at one time point only ($n=17$) or did not receive the VABS while participating in the RCTs.

Cognitive Functioning—Forty-nine children in the sample completed the Mullen Scales of Early Learning (MSEL; Mullen, 1995) at baseline. The MSEL yields standard scores for expressive language, receptive language, visual reception, fine motor skills, and an overall Verbal and Non-Verbal IQ Score (VIQ and NVIQ, respectively). Because, for some children, the child's age exceeded the standard cutoffs or their developmental level was too low, ratio IQs were calculated (see Bishop et al., 2011).

Data Analysis

Preliminary analyses—Based on the results of the initial BOSCC study (Grzadzinski et al., 2016), we aimed to establish a uniform distribution across the coding range (0–5) for the non-RRB items (Figure 3). Based on previous analyses (Grzadzinski et al., 2016; Kim & Lord, 2010), we did not expect to find a uniform distribution for the RRB items (Play, Sensory Interests, Hand/Finger Mannerisms, and Restricted/Repetitive Behaviors/Interests). Item distributions were averaged across segments A and B for the 13 ADOS-BOSCC items that make up the SC and RRB domains.

In order to confirm the factor structure of the ADOS-BOSCC, we conducted a confirmatory factor analysis using all Core items (SC and RRB; Table 2). Similar to the Standard BOSCC, we confirmed a two-factor model for the ADOS-BOSCC (SC and RRB) with the goodness-of-fit rating of a Comparative Fit Index (CFI) of 0.98 (CFI between 0.9 and 1 indicating good fit; Skrondal & Rabe-Hesketh, 2004) and a Root Mean Square Error Approximation (RMSEA) of 0.05 (RMSEA of 0.08 or less is considered a satisfactory fit; Browne & Cudeck, 1993). Notably, RRB items had lower factor loadings potentially due to the skewed distribution. See Table 2 below.

Primary statistical analyses

Inter-rater reliability: Twenty-two ADOS-BOSCC videos were coded by two or more coders for the purposes of assessing inter-rater reliability; two coders were chosen at random when scores were available from more than two coders. Two-way Random Absolute Intraclass Correlation Coefficients (ICCs) for inter-rater reliability were computed for the Core totals as well as the SC and RRB domains.

Test-retest reliability: A sub-sample of ADOS-BOSCC observations ($n=18$) from 9 children gathered about one month apart ($M=1.4$, $SD=0.47$) were coded to examine test-retest reliability. ICCs were calculated on the domain totals and individual item scores. Two-way Random Absolute Intraclass Correlation Coefficients (ICCs) for test-retest reliability were computed for the Core totals as well as the SC and RRB domains.

Validity: Changes in scores in the ADOS-BOSCC, Standard BOSCC and ADOS CSS from the first to the final observation were explored using paired T-tests. The magnitudes of changes were also examined using Cohen's *d* effect sizes. Consistent with previous work (Grzadzinski et al., 2016), individual growth change models were fitted to all the available data at multiple time points on each child for the ADOS-BOSCC SC and Core totals, Standard BOSCC SC and Core totals, and ADOS CSS SA and overall scores. For each participant, a linear regression was fitted and the coefficient associated with the age at assessment was used as the average rate of change score for that participant. We then standardized the expected change over 6 months by its standard deviation at baseline, which can be thought of as the effect size (Cohen's *D*) that would have been obtained using each measure had the children in the intervention been followed for 6 months from baseline and compared to a randomized control group showing no change. We used the 6 months duration to be consistent with the previous findings (Grzadzinski et al., 2016); however, the average time interval between T1 and T2 for our sample was closer to 9 months. Therefore, we repeated the same analysis with a 9-month-interval and found similar patterns of results (data available upon request). Additionally, correlations of cross-sectional and change scores were conducted across the ADOS-BOSCC, Standard BOSCC, and ADOS CSS scores to determine convergent validity. Finally, in order to control for the effects of socio-economic status of the children on the changes in scores to maximize discriminant validity and minimize coding contamination, we tested the effects of maternal education and race in a mixed model for repeated ADOS-BOSCC.

Post-hoc analyses: Following the process in Grzadzinski et al. (2016), responders and non-responders to treatment were identified based on changes between the first and last observations on the VABS, MSEL, and ADOS-2 CSS scores. Children who showed an increase in MSEL Receptive and/or Expressive Language scores of 5 points (1/2 a standard deviation) were classified as responders. Then, children were classified as responders if they showed an increase in 8 points (1/2 a standard deviation) on the VABS Communication Standard Score. Finally, children were classified as responders using the CSS if scores decreased 1 point (1 standard deviation). T-tests were conducted to compare change in the ADOS-BOSCC Core, SC, and RRB domains for responder and non-responder groups based on these measures.

Results

Inter-Rater Reliability

Based on the 22 videos coded by more than one coder, ICCs for ADOS-BOSCC Core, SC, and RRB were excellent, ranging from .88 to .96. Domain ICCs were .96 [95% CI (.85, .99)] for the Core, .93 [95% CI (.74, .98)] for the SC domain, and .88 [95% CI (.53, .97)] for the RRB domain (Supplement Table 1).

Test-Retest Reliability

Using a subset of children ($n=9$) with videos gathered about one month apart ($M=1.4$, $SD=.47$), ICCs for test-retest reliabilities were excellent, ranging from .85 to .88. Domain ICCs

were .87 [95% CI (.36, .97)] for the Core, .85 [95% CI (.26, .97)] for SC, and .85 [95% CI (.40, .97)] for the RRB domain (Supplement Table 2).

Validity

As shown in Figure 4, based on paired t tests, statistically significant decreases in scores (improvement in symptoms) were found in the ADOS-BOSCC scores from the first to the last observation for the SC domain ($M=5.48$, $SD=10.25$, $t(48)=3.74$, $p<0.05$; effect size = 0.6) and Core total ($M=-6.05$, $SD=12.65$, $t(48)=3.35$, $p<0.05$, effect size = 0.5). The ADOS-BOSCC RRB domain did not decrease significantly over time ($M=0.57$, $SD=4.44$, $t(48)=0.901$, $p=.37$, effect size = 0.2). Standard BOSCC scores (matched to ADOS-BOSCC observations) showed significant decreases in the SC domain ($M=-4.58$, $SD=7.4$, $t(48)=4.32$, $p<0.05$, effect size = 0.6), Core total ($M=6.19$, $SD=9.87$, $t(48)=4.39$, $p<0.05$, effect size = 0.6) and RRB scores ($M=-1.06$, $SD=3.53$, [$t(48)=2.10$, $p<0.05$, effect size = 0.3]. On the other hand, when ADOS CSS was used, no significant changes were noted for the CSS (CSS SA: $M=-0.51$, $SD=2.15$, [$t(48)=$, $p=0.10$, effect size = 0.25, CSS Total: $M=0.02$, $SD=2.02$, [$t(48)=$, $p=0.94$, effect size = -0.01]. ADOS CSS RRB showed significant increases over time ($M=0.918$, $SD=1.79$, $t(48)=-3.59$, $p<0.05$, effect size = -0.50).

The average rates of change in the ADOS-BOSCC SC and Core totals over 6 months were moderate (Cohen's $d = 0.3$) based on individual growth models accounting for all time points. The average rates of change in the Standard BOSCC SC and Core totals over 6 months were also moderate (Cohen's $d = 0.5$). When we repeated the same analysis with a 9-month-interval, we found similar patterns of results with larger effect sizes ranging from 0.4 to 0.8 (data available upon request). The corresponding values for ADOS CSS SA domain and overall CSS were small (Cohen's $d = 0.1$).

Cross sectional correlations revealed that ADOS-BOSCC Core and Standard BOSCC Core totals were strongly correlated ($r=.80$, $p<0.01$). ADOS-BOSCC Core and ADOS CSS were also significantly correlated ($r=.32$, $p<0.01$). Change scores in ADOS-BOSCC Core totals from Time 1 to Time 2 were also strongly correlated with change scores in the Standard BOSCC Core totals ($r=.45$, $p<0.01$). Change scores for the ADOS-BOSCC Core and change scores in the ADOS CSS were not significantly correlated ($r=.15$, $p=.30$).

Using mixed models, we confirmed that maternal education, gender and race were not significantly related to the changes in ADOS-BOSCC SC (maternal education, $F=0.277$, $p=0.896$; gender $F=2.062$; $p=0.153$; race, $F=0.128$, $p=0.15$), RRB (maternal education, $F=0.547$, $p=0.70$; gender, $F=-.254$, $p=0.615$; race, $F=0.041$, $p=0.83$) or Core (maternal education, $F=0.213$, $p=0.93$; gender, $F=1.722$, $p=0.19$; race, $F=0.119$, $p=0.73$) totals.

Post-hoc Analyses

Based on paired T-tests, significant decreases between time points for the ADOS-BOSCC SC and Core totals were observed for the responders based on the MSEL Receptive (SC; ($M=8.9$, $SD=9.65$), [$t(21)=4.34$, $p<.001$], Core; ($M=9.9$, $SD=12.39$), [$t(21)=3.74$, $p<.01$]), MSEL Expressive (SC; ($M=9.8$, $SD=8.2$), [$t(14)=4.6$, $p<.001$], Core; ($M=10.7$, $SD=10.2$),

[$t(14)=4.0, p<.01$]), and VABS Communication (SC; ($M=9.5, SD=10.0$), [$t(16)=3.9, p<.01$], Core; ($M=10.7, SD=12.3$), [$t(16)=3.6, p<.01$]) domain scores.

Based on change scores between first and last time point, T-tests revealed significant differences in the ADOS-BOSCC SC and Core totals between Responders and Non-Responders based on the MSEL Receptive (SC; ($M=7.2, SEM=2.9$), [$t(36)=2.5, p<.05$], Core; ($M=7.9, SEM=3.7$), [$t(36)=2.1, p<.05$]) and Expressive (SC; ($M=7.9, SEM=2.9$), [$t(33)=2.7, p<.05$], Core; ($M=8.1, SEM=3.9$), [$t(33)=2.1, p<.05$]) domain scores. Additionally, T-tests indicated differences in the ADOS-BOSCC Core between the Responders and Non-Responders based on the VABS Communication domain (Core; ($M=8.2, SEM=3.8$), [$t(30)=2.1, p<.05$]).

Discussion

The results of the study indicate that the Standard BOSCC coding scheme can be applied to selected videotaped segments from the ADOS (ADOS-BOSCC) to measure subtle changes in social communication over time in young, minimally verbal children with ASD. Similar to the results from the Standard BOSCC when applied to parent-child play interactions (Grzadzinski et al., 2016) and the symptom domains operationalized under the DSM-5 (American Psychiatric Association; 2013), we confirmed that the ADOS-BOSCC items are clustered under two different factors, SC and RRB. This allows researchers and clinicians to monitor changes in ASD symptoms separately for the SC and RRB domains. Consistent with previous findings of the Standard BOSCC (Grzadzinski et al., 2016), we have demonstrated consistent decreases in symptom levels in the SC domain with the ADOS-BOSCC and Standard BOSCC scores, whereas the results based on the RRBs were more varied. These results also indicate that the ADOS-BOSCC has excellent inter-rater and test-retest reliability.

It is encouraging that the changes measured by the ADOS-BOSCC and Standard BOSCC were fairly comparable to each other. First, ADOS-BOSCC and Standard BOSCC scores were strongly correlated with each other. Moreover, significant decreases in symptom levels based on the ADOS-BOSCC were primarily seen on the SC domain scores as well as the Core totals, which combines the SC and RRB domain scores, aligning with previous work based on Standard BOSCC (Grzadzinski et al., 2016). The effect sizes of change observed from the Standard BOSCC and ADOS-BOSCC ranged from small to moderate (ranging 0.3–0.6), and the effect sizes based on the Standard BOSCC in these areas were either comparable to, or slightly larger than those based on the ADOS-BOSCC, ranging from 0.5–0.6. The consistencies in the patterns of changes we observed between the ADOSBOSCC and the Standard BOSCC are especially encouraging given that these behavioral changes were rated based on different contexts. More specifically, the ADOS-BOSCC was rated based on examiner-child interactions whereas the Standard BOSCC was rated based on parent-child interactions. These results are consistent with previous work suggesting that the BOSCC, when applied to parent-child play segments or ADOS segments, may be more sensitive than the ADOS CSS in capturing subtle changes in social communication symptoms in response to short-term treatment (Grzadzinski et al., 2016).

This work also confirmed convergent validity of the ADOS-BOSCC in detecting behavioral changes measured by other instruments, including the MSEL and VABS-2. Children who were considered “Responders” to treatments based on the MSEL (Receptive and Expressive Language domains) and VABS (Communication Domain) showed significant changes in SC domain and Core total of the ADOS-BOSCC. These results confirm previous work (Grzadzinski et al., 2016) and suggest that the ADOS-BOSCC can successfully capture changes in core symptoms of ASD that are in line with clinically meaningful changes in parent-reported adaptive skills and standard developmental testing. Notably, since the ADOS segments were selected in a standardized fashion, the validity of the application of the Standard BOSCC on other segments of the ADOS is still unclear. In addition, ADOS administrations were conducted by research reliable administrators who were blind to treatment status. Therefore, the application of the Standard BOSCC to ADOS administrations conducted by clinicians who have not achieved research reliability and/or are not blind to the child’s treatment status are unknown.

These results indicate that changes in the RRB domain may vary depending on the method used to identify those behaviors. The ADOS-BOSCC did not capture any significant changes in RRBs over time even though we saw a trend for a decrease in scores; however, changes in RRBs were minimal, as evidenced by the small effect size of 0.1. The Standard BOSCC RRB domain demonstrated significant decreases over time for our sample, even though the initial finding with the Standard BOSCC did not show any significant changes in RRBs (Grzadzinski et al., 2016). The average ADOS-BOSCC RRB domain scores were higher than the average Standard BOSCC RRB scores at both T1 and T2. The reason why we observed more severe or frequent RRBs and less changes in RRBs based on the ADOS-BOSCC compared to the Standard BOSCC may be partly because the ADOS includes standardized tasks (e.g., bubble and balloon plays) that are designed to elicit RRBs in young children. Therefore, the ADOS-BOSCC was also coded based on 12-minute examiner-child interactions, which may provide more opportunities to observe RRBs, whereas the Standard BOSCC was based on 10-minute parent-child interactions. In contrast, ADOS CSS RRB showed significant increases over the same period of time. This may be partly because the ADOS CSS is fairly independent of developmental factors such as the child’s age and language. Therefore, the severity of the RRBs by the ADOS CSS for one child is measured in comparison with other children of similar age and language levels; as the child ages and his/her language progresses from T1 to T2, the severity of RRBs may be measured in comparison with different age and language groups, unlike the ADOS-BOSCC scores. The ADOS CSS RRB domain also includes behaviors such as intonation and stereotyped language which may increase from T1 to T2 as children’s language skills develop over time. Also, play skills are only captured by the ADOS-BOSCC, not by the ADOS CSS RRB domain. Finally, the ADOS-BOSCC is based on 12-minute parent-child interaction videos whereas the ADOS CSS is based on 40- to 60-minute examiner-child observations, which may provide more opportunities to observe a wider range of RRBs. However, the ADOS-BOSCC observations were rated by coders unaware of treatment status and time points, while the ADOS CSS scores were given by expert clinicians who were unaware of the child’s treatment stat, but not time point, which may introduce some bias. These results as a whole may suggest that, in the absence of standard opportunities created intentionally to

observe RRBs such as during certain ADOS tasks (e.g., bubble and balloon plays), the Standard BOSCC scores based on parent-child interactions may be able to detect decreases in RRBs in young children who receive treatment. However, these changes need to be interpreted cautiously while considering the impact of maturation and language development on the BOSCC scores, which can be addressed by having a control group in RCT designs.

Limitations and Future Directions

Although the results are promising as initial evidence for the validity of the application of the Standard BOSCC coding to video-recorded ADOS segments, several limitations should be considered. The results of this study were based on a relatively small sample size ($n=49$) of children who were minimally-verbal and under age 5. Replication of these results and extensions to older children remains to be explored. A subset of the sample ($n=32$) overlapped with the sample of children who were included in the previous study examining psychometrics of the Standard BOSCC (Grzadzinski et al., 2016), highlighting the need for replications in new samples. Furthermore, since the focus of this study was to examine the validity and reliability of the ADOS-BOSCC, we did not test the specific effects of treatment nor compare the effects of different types of treatment on behavioral changes measured by the ADOS-BOSCC, which need to be explored more in depth in future studies. Replications using the ADOS-BOSCC with other larger, more representative, independent samples will inform the validity of the measure in other populations before it can be generalized in other research and clinical settings. Finally, the initial development of the BOSCC was focused on minimally verbal children given the importance and need for the outcome measure to evaluate the effectiveness of early intervention. However, because the manifestation of ASD symptoms and target behaviors of early interventions vary by developmental levels, the development of coding schemes that are more appropriate for children with flexible and complex speech is currently underway.

Conclusion

The goal of the study was to determine the validity and reliability of applying the Standard BOSCC, a newly validated treatment outcome measure, to standardized selections of videotaped ADOS segments (ADOS-BOSCC). The results suggest that the ADOS-BOSCC is more sensitive to monitoring changes over the course of treatment compared to the ADOS CSS and thus can be applied to pre-existing datasets of ADOS videos to examine treatment response. This study provides support for the utility of the ADOS-BOSCC in identifying subtle changes in social communication over short periods of time in young, minimally-verbal children with ASD. Additional studies with new samples are warranted in order to elucidate the benefits and limitations of the ADOS-BOSCC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This work was supported by grants awarded to C.L. from the NIMH (R01MH081757, 1RC1MH089721, R01RFAMH14100, R01MH078165), Autism Speaks (5766), and HRSA (UA3MC11055). This work was also

supported by a Dennis Weatherstone Predoctoral Fellowship from Autism Speaks and a Graduate Student fellowship with Weill Cornell Medical College and Teachers College, Columbia University awarded to author R.G. We thank children and families who participated in the study. We also thank Nurit Benrey, Yeo-Bi Choi, Morgan Cohen, Michelle Heyman, Allison Megale, Gabrielle Gunnin, Gabrielle Ranger-Murdock, Anna Marie Paolicelli, and Anya Ucruyo for assistance with coding and Sheri Stegall at Western Psychological Services for copyright assistance.

References

- Anagnostou E, Jones N, Huerta M, Halladay AK, Wang P, Scahill L, ... Dawson G (2015). Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism: The International Journal of Research and Practice*, 19(5), 622–636. 10.1177/1362361314542955 [PubMed: 25096930]
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Bishop SL, Guthrie W, Coffing M, & Lord C (2011). Convergent validity of the Mullen Scales of Early Learning and the differential ability scales in children with autism spectrum disorders. *American journal on intellectual and developmental disabilities*, 116(5), 331–343. [PubMed: 21905802]
- Bolte EE, & Diehl JJ (2013). Measurement tools and target symptoms/skills used to assess treatment response for individuals with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 43(11), 2491–2501. 10.1007/s10803-013-1798-7 [PubMed: 23479074]
- Brian J, Smith I, Zwaigenbaum L, Roberts W, & Bryson S (2015). The social ABCs caregiver-mediated intervention for toddlers with autism spectrum disorder: feasibility, acceptability, and evidence of promise from a multisite study. *Autism Research*, doi: 10.1002/aur.1582
- Browne MW, & Cudeck R (1993). Alternative ways of assessing model fit In Bollen KA & Long JS (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Cunningham AB (2012). Measuring change in social interaction skills of young children with autism. *Journal of Autism and Developmental Disorders*, 42(4), 593–605. [PubMed: 21638109]
- Dawson G, Rogers S, Munson J, Smith M, Winter J, Greenson J, et al. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics*, 125(1), e17–e23. doi: 10.1542/peds.2009-0958 [PubMed: 19948568]
- Fletcher-Watson S, Petrou A, Scott-Barrett J, Dicks P, Graham C, O'Hare A, ... & McConachie H (2016). A trial of an iPad™ intervention targeting social communication skills in children with autism. *Autism*, 20(7), 771–782. [PubMed: 26503990]
- Gotham K, Pickles A, & Lord C (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(5), 693–705. [PubMed: 19082876]
- Green J, Charman T, McConachie H, Aldred C, Slonims V, Howlin P, ... Pickles A (2010). Parent-mediated communication-focused treatment in children with autism (PACT): a randomised controlled trial. *The Lancet*, 375(9732), 2152–2160. 10.1016/S0140-6736(10)60587-9
- Grzadzinski R, Carr T, Colombi C, McGuire K, Dufek S, Pickles A, & Lord C (2016). Measuring Changes in Social Communication Behaviors: Preliminary Development of the Brief Observation of Social Communication Change (BOSCC). *Journal of Autism and Developmental Disorders*, 46(7), 2464–2479. 10.1007/s10803-016-2782-9 [PubMed: 27062034]
- Hus V, Gotham K, & Lord C (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*, 44(10), 2400–2412. [PubMed: 23143131]
- Kasari C, Lawton K, Shih W, Barker TV, Landa R, Lord C, ... Senturk D (2014). Caregiver-Mediated Intervention for Low-Resourced Preschoolers With Autism: An RCT. *Pediatrics*, 134(1), e72–e79. 10.1542/peds.2013-3229 [PubMed: 24958585]
- Kim SH, & Lord C (2010). Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the Autism Diagnostic Observation Schedule (ADOS). *Autism Research*, 3(4), 162–173. [PubMed: 20589716]

- Kitzerow J, Teufel K, Wilker C, & Freitag CM (2016). Using the brief observation of social communication change (BOSCC) to measure autism-specific development. *Autism Research*, 9(9), 940–950. 10.1002/aur.1588 [PubMed: 26643669]
- Lord C, Rutter M, & Couteur A (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659–685. doi: 10.1007/BF02172145 [PubMed: 7814313]
- Lord C, Rutter M, DiLavore PC, Risi S, Gotham K, & Bishop SL (2012). *Autism Diagnostic Observation Schedule, second edition (ADOS-2) modules 1–4*. Los Angeles, CA: Western Psychological Services.
- Matson JL (2007). Determining treatment outcome in early intervention programs for autism spectrum disorders: A critical analysis of measurement issues in learning based interventions. *Research in Developmental Disabilities*, 28(2), 207–218. 10.1016/j.ridd.2005.07.006 [PubMed: 16682171]
- Mullen EM (1995). *Mullen scales of early learning*. Circle Pines, MN: American Guidance Service.
- Pijl MK, Rommelse NN, Hendriks M, De Korte MW, Buitelaar JK, & Oosterling IJ (2016). Does the Brief Observation of Social Communication Change help moving forward in measuring change in early autism intervention studies? *Autism*, 1362361316669235. 10.1177/1362361316669235
- Rogers SJ, Estes A, Lord C, Vismara L, Winter J, Fitzpatrick A, ... Dawson G (2012). Effects of a brief Early Start Denver model (ESDM)-based parent intervention on toddlers at risk for autism spectrum disorders: a randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51(10), 1052–1065. 10.1016/j.jaac.2012.08.003 [PubMed: 23021480]
- Skrondal A, & Rabe-Hesketh S (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Crc Press.
- Shumway S, Farmer C, Thurm A, Joseph L, Black D, & Golden C (2012). The ADOS calibrated severity score: Relationship to phenotypic variables and stability over time. *Autism Research*, 5(4), 267–276. doi: 10.1002/aur.1238 [PubMed: 22628087]
- Sparrow SS, Cicchetti DV, & Balla DA (2005). *Vineland adaptive behavior scales, (Vineland-II)*. Circle Pines, MN: American Guidance Services.
- Thurm A, Manwaring SS, Swineford L, & Farmer C (2015). Longitudinal study of symptom severity and language in minimally verbal children with autism. *Journal of Child Psychology and Psychiatry*, 56(1), 97–104. doi: 10.1111/jcpp.12285 [PubMed: 24961159]
- Wetherby AM, Guthrie W, Woods J, Schatschneider C, Holland RD, Morgan L, & Lord C (2014). Parent-Implemented Social Intervention for Toddlers With Autism: An RCT. *Pediatrics*, 134(6), 1084–1093. 10.1542/peds.2014-0757 [PubMed: 25367544]

	Item	Domain	
1	Eye Contact	Social-Communication (maximum score: 45**)	Core (maximum score: 65***)
2	Facial Expressions		
3	Gestures		
4	Vocalizations		
5	Integration of Vocal and Non-Vocal		
6	Social Overtures		
7	Social Responses		
8	Requesting*		
9	Engagement		
10	Play	RRB (maximum score: 20)	
11	Unusual Sensory Interests		
12	Hand/Finger/Body Mannerisms		
13	Repetitive/Stereotyped Interests/Behaviors	Other Abnormal Behaviors (maximum score: 15)	
14	Activity Level		
15	Disruptive Behavior/Irritability		
16	Anxious Behaviors		

RRB Restricted, repetitive behavior/interest

*Requesting is unique to ADOS-BOSCC

**Standard BOSCC maximum Social-Communication score: 40

***Standard BOSCC maximum Core score: 60

Figure 1.
 ADOS-BOSCC items, domains, and total

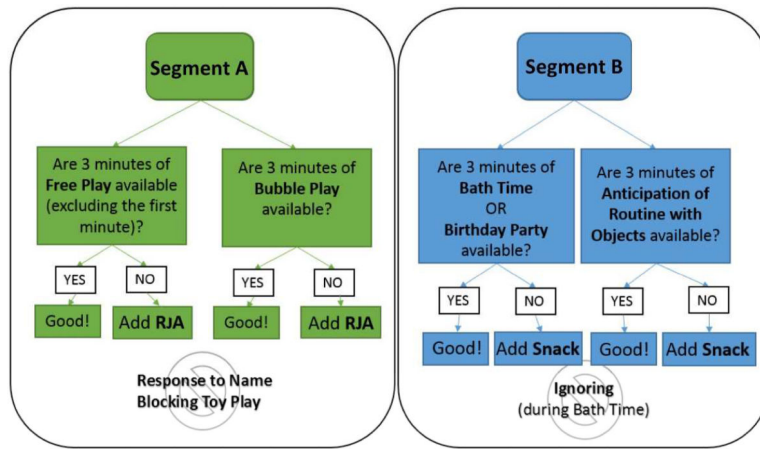


Figure 2.
Selection of ADOS-BOSCC segments

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Distributions for 13 core ADOS-BOSCC items (averaged across segments A & B). Solid red represents social communication items, while the striped blue represents items in the restricted, repetitive behavior domain

Figure 3.
ADOS-BOSCC item distributions

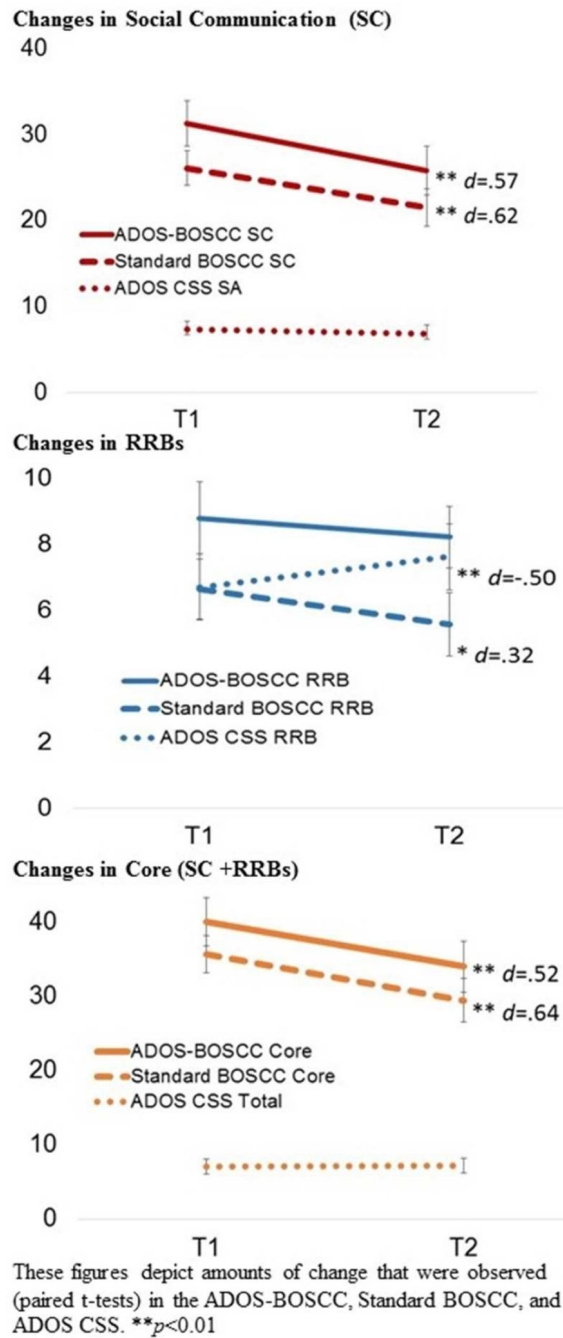


Figure 4. Changes in Social Communication (SC), RRBs and Core (SC + RRB) scores by measure

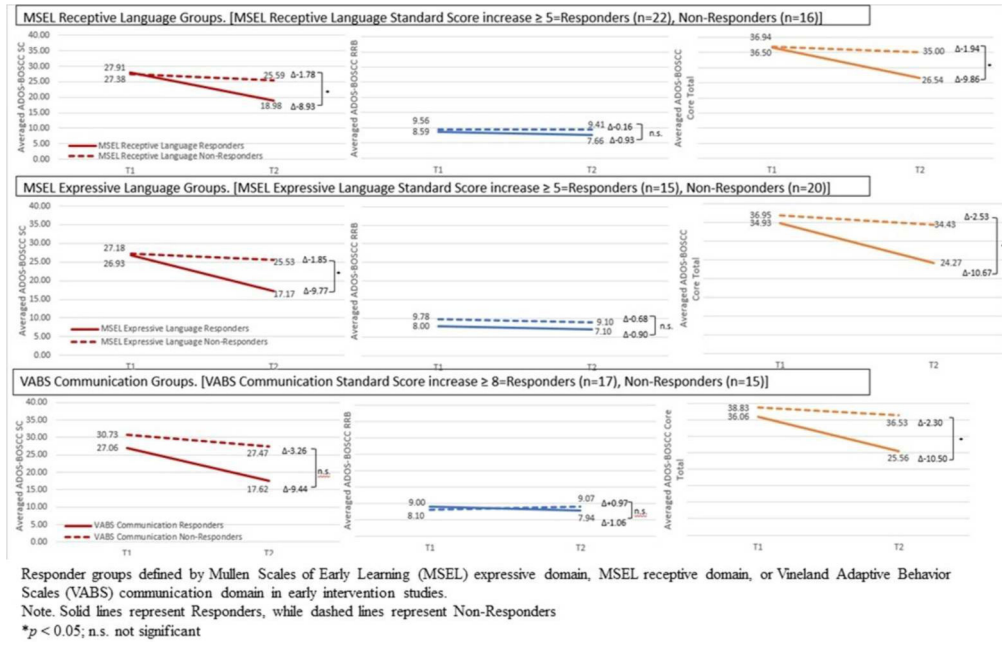


Figure 5.
 Responder groups defined by MSEL and VABS

Table 1.Demographic and Baseline Characteristics ($N=49$)

	Mean (SD)	Range
Age (months)	23.2 (9.5)	16–55
VABS (standard score) ($n=32$)		
Communication	76.2 (16.1)	47–121
Socialization	85.2 (10.5)	68–110
Daily living	86.5 (12.9)	65–113
Motor skills	91.2 (12.9)	69–129
MSEL (ratio) ($n=49$)		
VIQ	60.1 (24.7)	23–145
NVIQ	82.9 (20.8)	36–132
ADOS-2 ($n= 49$)		
CSS	7.1 (2.0)	1–10
SA CSS	7.3 (2.1)	1–10
RRB CSS	6.7 (1.7)	5–10
		n (%)
Sex (males)		39 (80)
Race ^a		
Caucasian		36 (74)
African American		4 (8)
Other		7 (14)
Ethnicity ^b (Hispanic)		1 (2)
Maternal education ^c (4+ years of college)		27 (55)

ADOS-2 Autism Diagnostic Observation Schedule, 2nd Edition, *CSS* Calibrated Severity Score, *MSEL* Mullen Scales of Early Learning, *RRB* *CSS* Restricted, Repetitive Behavior Calibrated Severity Score, *SA CSS* SocialAffect Calibrated Severity Score, *SD* standard deviation, *VABS* Vineland Adaptive Behavior Scales

^a Two participants (4%) did not report race information

^b Two participants (4%) did not report ethnicity information

^c One participant (2%) did not report information about maternal education

Table 2.

Confirmatory factor analyses (CFA) for ADOS-BOSCC

2-Factor model			
Factor 1		Factor 2	
Eye Contact	0.67	Play	0.75
Facial Expressions	0.61	Unusual sensory interests	0.51
Gestures	0.73	Hand/finger/body mannerisms	0.10
Vocalizations	0.85	Repetitive interests/behaviors	0.37
Integration of vocal and non-vocal	0.90		
Social overtures	0.82		
Social responses	0.71		
Requests	0.84		
Engagement	0.73		

2-factor model loadings for the ADOS-BOSCC items; all factor loadings ≥ 0.4 shown in bold