Measuring child social communication across contexts: Similarities and differences across play and snack routines.

Kyle M. Frost, Genna N. Koehn, Kaylin M. Russell, and Brooke Ingersoll

Michigan State University

Acknowledgements

This project was supported by the Congressionally Directed Medical Research Programs under grant number W81XWH-10-1-0586, and the Health Resources and Services Administration (HRSA) of the U.S. Department of Health and Human Services (HHS) under grant number R40MC27704, and the R40 Maternal and Child Health Field-initiated Innovative Research Studies Program. This information or content and conclusions are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS or the U.S. Government.

The last author receives royalties from the sale of the manual used in the research. Royalties are donated to the research. We thank the families who participated in this study.

Lay Summary

Improving measurement of outcomes in studies of early interventions for autism spectrum disorder (ASD) has been identified as a priority in the field. The importance of measurement across contexts has also been indicated by researchers and community stakeholders. The goal of this study was to determine whether an existing observational rating scheme, the Brief Observation of Social Communication Change (BOSCC), could be applied to a new activity: a home snack routine. Results suggest that rating the BOSCC on home snack routines is feasible and promising for capturing change over time. In addition, some child behaviors differed across play and snack, lending further support for the claim that measurement across activities is essential.

Abstract

Improving measurement of outcomes in randomized controlled trials of early interventions for autism spectrum disorder (ASD) has been identified as a priority in the field. In addition, the importance of measurement across contexts has been indicated by researchers and community stakeholders alike (Lord et al., 2005; McConachie et al., 2015; Schreibman et al., 2015). The Brief Observation of Social Communication Change (BOSCC; Grzadzinski et al., 2016), an observational rating scheme of brief play interactions, was developed to address a need for measures that are reliable, sensitive to change, and valid for use in research settings. The goal of this study was to examine the feasibility and utility of applying the BOSCC to a new context: a home snack routine. Results suggest that rating the BOSCC on home snack routines is feasible and psychometrically sound, and captures change in child social communication behaviors. However, the utility of the BOSCC for measuring restricted and repetitive behaviors (RRBs) is less clear. Nonetheless, differences in RRBs across play and snack lend support for the claim that measurement across contexts is essential. Application of the BOSCC across contexts may allow researchers to obtain a more accurate estimate of intervention response and help capture context-specific changes in social communication. It may also provide a method for researchers to evaluate the effect of context on child social communication more broadly.

Key Words: Social communication, measurement, ASD

Measuring child social communication across contexts: Similarities and differences across play and snack routines.

To address a growing need for high-quality interventions for autism spectrum disorder (ASD), numerous randomized controlled trials of early interventions for ASD have been completed, most of which have been published in the last decade (French & Kennedy, 2018). Current best practices in treatment of ASD in young children involve an integration of developmental and behavioral approaches, and inclusion of caregivers in their child's treatment (National Research Council, 2001; Zwaigenbaum et al., 2015). Several evidence-based interventions have been developed for or adapted to caregiver-mediated models, in which caregivers are trained to provide intervention to their children during play and daily activities (reviewed in Nevill, Lecavalier, & Stratis, 2016; Oono, Honey, & McConachie, 2013). This shift toward more naturalistic treatments and increases in caregiver involvement reflect the importance of teaching skills in a way that generalizes across people and contexts (Schreibman et al., 2015; Vismara & Rogers, 2010).

Despite the importance of evaluating treatment outcomes, no consensus or gold-standard measure has emerged in the field (Anagnostou et al., 2015). Bolte & Diehl (2013) reviewed 195 studies, and found that over 200 outcome measures were used to evaluate response to intervention, with the majority of these measures used in a single study. Likewise, a recent review of randomized controlled trials examined 48 studies, which reported 87 different outcome measures (French & Kennedy, 2018). These measures comprise a variety of methods, including caregiver-report questionnaires, clinician ratings, and direct observation. They also represent varying degrees of proximity to treatment targets, including measures of proximal outcomes (i.e. those that directly relate to treatment targets) as well as more distal outcomes (i.e. those which

are not explicit targets of treatment). Heterogeneity in measurement approaches is a barrier to understanding the efficacy of early interventions for ASD, as it inhibits comparison across studies.

The Brief Observation of Social Communication Change (BOSCC; Grzadzinski et al., 2016) was developed to address current limitations to measuring outcomes in early intervention trials for ASD. The BOSCC is an observational rating scheme which measures social communication behaviors during brief play interactions between children and their caregivers or other adult play partners. Items in the BOSCC were adapted from the second edition of the Autism Diagnostic Observation Schedule (ADOS-2; Lord et al., 2012), with an expanded coding range to increase sensitivity to change. Based on a validation by the measure authors, the BOSCC appears to be reliable across raters and brief time intervals and shows convergent validity with other measures of social communication (Grzadzinski et al., 2016). Results in terms of sensitivity to change have been mixed, with some studies finding that the BOSCC captures change (Grzadzinski et al., 2016; Kitzerow, Teufel, Wilker, & Freitag, 2016; Pijl et al., 2016), and others finding no significant change over time (Fletcher-Watson et al., 2016; Nordahl-Hansen, Fletcher-Watson, McConachie, & Kaale, 2016). Whether this variation is due to inadequate sensitivity or lack of treatment effects in a subset of studies is unclear.

To understand treatment efficacy, it is important to use measures that address meaningful change and evaluate generalization of skills across contexts (Lord et al., 2005; Schreibman et al., 2015). Measurement across contexts is useful for several reasons. First, research suggests that child social communication behaviors such as eye contact vary across activities (Jones et al., 2017); therefore, measurement in more than one type of interaction may lead to more accurate estimates of ability. Additionally, caregiver behaviors such as language use differ across contexts

in typical development. In research with mother-infant dyads, mothers' language use fluctuated during daily routines (with periods of silence interspersed with periods of talking), while language use during play was more stable throughout a play interaction (Tamis-LeMonda, Custode, Kuchirko, Escobar, & Lo, 2018; Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017). Thus, measurement across contexts may better account for natural variation in child and caregiver communication in different daily activities. In addition to being indicated by research, measurement across contexts is supported by community stakeholders. Individuals on the autism spectrum have recommended direct observation in multiple settings; likewise, caregivers of autistic children also recommended using videos of relaxed interactions in multiple contexts to measure change over time (McConachie et al., 2015).

In addition to contributing to the growing body of evidence characterizing the utility of the BOSCC for measuring change in play interactions, this study extends the literature by examining the feasibility and utility of applying the BOSCC to a new context: a snack routine. The ability to apply the BOSCC across contexts has implications for researchers' ability to measure the generalization of social communication skills across home contexts, to evaluate the efficacy of interventions targeting behavior during daily routines, and to reliably estimate children's functional skills.

Method

Participants

Participants were 50 children and their caregivers who participated in two research studies evaluating the efficacy of a low-dose parent-mediated intervention delivered via telehealth. These two studies included a completed pilot randomized-controlled trial (RCT; Ingersoll, Wainer, Berger, Pickard, & Bonter, 2016), and an ongoing full-scale RCT. Both

research studies obtained approval from the Human Research Protection Program at Michigan State University, and all families provided informed consent to participate in this research. Participant demographics and descriptive statistics can be found in Table 1.

[Insert Table 1. about here]

Intervention

All caregivers in this dataset received access to an online adaptation of Project ImPACT (Ingersoll & Dvortcsak, 2009; Ingersoll et al., 2016), either at a self-directed pace or with the support of a coach on a biweekly basis. Project ImPACT is a caregiver-mediated intervention, which provides caregivers with skills to improve their children's social communication skills, including social engagement, imitation, play, and communication. Caregivers are gradually exposed to the intervention strategies in the period between pre- and post-treatment assessments. Research indicates that caregivers improve in their use of strategies during this time, and fidelity to the treatment is maintained between post-treatment and follow-up (Ingersoll et al., 2016).

Measures

BOSCC

This study used a pre-publication research version of the BOSCC (version 8/17/16) with permission of WPS and the measure authors. This measure was used to rate videos of caregiver-child interactions across three timepoints and two contexts. Primary raters were kept blind to both child treatment status (i.e. self-directed or therapist-assisted) and video time point when coding the BOSCC. The BOSCC measure consists of 15 items, which rate child social communication behaviors and the presence of restricted and repetitive interests and behaviors. Each item is rated on a six-point scale from 0-5, with higher scores representing greater impairment in these domains. Items 1-8 comprise the Social Communication (SC) domain and

items 9-12 comprise the Restricted and Repetitive Behavior (RRB) domain, with all 12 items comprising the Core Total score. Items 13-15, which reflect other behaviors which may interfere with the BOSCC administration, were rarely coded and are therefore not included in analyses.

Videos were rated in two five-minute segments, with final BOSCC scores based on the average summed score across the two segments. Videos less than 10 minutes long but greater than 8 minutes long were split into two halves of equal length for rating. When videos were longer than 10 minutes, only minutes 0:00 to 10:00 were coded. Some shorter videos were recorded (i.e. less than 8 minutes long; n=59), in which case the BOSCC rating scheme was applied to a single five-minute segment (minutes 0:00 to 5:00). An independent samples *t*-test was used to compare scores for the 5-minute and 10-minute videos. Core Total scores did not differ, $t(328) = -0.52$, $p = 0.60$; therefore, all available scores were included in the main analyses.

Other child characteristics

At Time 1, The Autism Diagnostic Observation Schedule, 2nd Edition (ADOS; Lord et al., 2012) was used to confirm child eligibility for both research studies. In addition, the Mullen Scales of Early Learning (MSEL; Mullen, 1995) and the Vineland Adaptive Behavior Scales, 2nd edition (VABS; Sparrow, Cicchetti, & Balla, 2005) were administered at Time 1 to evaluate developmental level and adaptive functioning.

Caregiver-child interactions: Play and Snack videos

Caregiver-child interactions were filmed in participants' homes at three time-points. Time 1 took place before caregivers accessed the intervention, shortly after intake assessments. Time 2 occurred immediately post-intervention, 4-6 months after intake assessments. Time 3 took place 3- to 4-months after Time 2. Videos included a 10-minute free-play interaction with a set of toys provided by the study, as well as a 10-minute snack routine. For the play routine,

caregivers were instructed to play as they normally would with their child, using as many or as few of the toys in the box as they preferred. For the snack routine, caregivers were told that we wanted to capture all the parts of the family's usual home snack routine from the beginning, including things like preparing or selecting a snack and cleaning up if that was what they normally did at home.

Training Process

The first author (KF) trained to reliability with the authors of the BOSCC instrument. Two undergraduate research assistants (GK and KR) were then trained to reliability to the first author. After two 2-hour meetings on didactics (e.g. defining terms, explaining coding protocols, reviewing the contents of each code), all three coders rated and discussed videos until trainees were reliable. To be considered reliable, trainees were required to be within 1 point on at least 80% of items, and within 4 points of the total score, on both segments of three consecutive videos (i.e. six consecutive segments). Undergraduate coders achieved reliability upon completion of 15 and 17 videos, coded over the course of one month. After reaching reliability, the group continued regular consensus-coding discussions to prevent drift (approximately weekly).

Adapting the BOSCC for snack routines

The BOSCC was developed for rating play interactions between a child and adult. It was adapted for rating snack routines for the present study, with permission from the measure authors (personal communication, C. Lord & R. Grzadzinski, January, 2017). Most items were not altered for rating snack routines. Broadly, any references to "play interaction" or "play activity" were considered to refer to the interaction or activity more generally. The following adaptations to the rating scheme were made for rating snack routines:

- Redefining "Activity": Several items refer to different activities, defined in the BOSCC Glossary as "various discrete events or play tasks that occur during the session." Typically, in a play session, different activities are comprised of play with different toys, or different types of play with the same toy set. Before rating videos, coders met and discussed what would be considered different activities in the context of a snack routine, including things like food preparation, selecting a snack, eating or drinking, clean up, and washing hands. Eating different foods, or having a snack and then a drink, were *not* considered two different activities.
- Engagement: References to play were removed from the language of the item "Engagement in Play Activity/Interaction" (#8), and the item was reworded to refer to engagement in the interaction in more generic terms.
- Play: The "Play with Objects" item (#9) was not coded when rating snack routines.

In the context of snack routines, families occasionally played while having a snack, or transitioned to play after the child finished eating but before 10 minutes had elapsed. There was no statistically significant difference in BOSCC SC domain or Core Total at any time point between children who had some play intermixed with their snack routine, and those who did not (p's > 0.34). Therefore, for all analyses, BOSCC ratings on the entire snack routine interaction were used, even if some play occurred while children were eating snack.

Analysis Plan

The goals of this study were to determine the feasibility and utility of adapting the BOSCC instrument for use in rating home routines. Therefore, psychometric properties of BOSCC Core Total, SC domain, and RRB domain scores were evaluated for both play routines

(BOSCC-Play) and snack routines (BOSCC-Snack); this included internal consistency, inter-rater reliability, and inspection of item distributions.

Next, we were interested in understanding the relationship between BOSCC domain and Core Total scores across contexts. Correlations between BOSCC-Play and BOSCC-Snack domain and Core Total scores were used to evaluate the concurrent relationship between BOSCC scores in the two contexts. We expected that these scores would be significantly correlated across contexts. To understand whether there were mean differences across contexts, paired-samples *t*-tests were used to evaluate whether the BOSCC-Play and BOSCC-Snack domain and Core Total scores differed across contexts. We expected that the measure would capture similar levels of social communication across contexts (i.e. the difference would not be significant). However, given previous research demonstrating variability in RRB presentation across contexts (Stronach & Wetherby, 2014), we were uncertain about whether this would hold true for the RRB domain.

To understand the utility of the BOSCC-Snack for use in intervention research, a multilevel linear model was used to clarify if and how Core Totals for the BOSCC-Play and BOSCC-Snack changed over the course of a caregiver-mediated intervention trial. Given that the BOSCC has been conceptualized as a measure sensitive to short-term change, we expected to see significant improvement in BOSCC scores over time, including change from pre- to post-intervention, and change from post-intervention to follow-up.

To better understand the incremental validity of adapting the BOSCC scheme to a second context, hierarchical regressions were used to understand the unique predictive variance added by the BOSCC-Snack Core Total score for relevant outcome measures. In other words, we were interested in understanding whether measuring child social communication in two contexts

provided a better estimate of other child characteristics (e.g. developmental level and functional communication skills) than a single context.

Results

Scale Reliability

Our first goal was to replicate the psychometric properties of the BOSCC-Play and explore the psychometric properties of the BOSCC-Snack. BOSCC Core Total and SC domain scores were adequately normally distributed for both BOSCC-Play and BOSCC-Snack (Table 2; Figure 1). Consistent with results from Grzadzinski et al. (2016), RRB domain scores were positively skewed for both contexts.

[Insert Figure 1. about here]

Cronbach's alpha was used to evaluate the internal consistency of the items within the SC domain, RRB domain, and Core Total scores. The BOSCC-Play and BOSCC-Snack both demonstrated fair to good internal consistency for the SC domain and for the Core Total (Table 2); however, internal consistency was poor for BOSCC-Play RRB domain score (Cicchetti, 1994).

In addition, a total of 114 videos (63 play routines, 51 snack routines) were rated independently by two people. For each child, one play video and one snack video were selected at random for reliability coding. Of those randomly selected, 44 were at Time 1, 40 were at Time 2, and 30 were at Time 3. Intra-class correlation (ICC) was used to evaluate inter-rater reliability. A single-measures, two-way mixed design based on absolute agreement was used (Table 2). Inter-rater reliability was excellent across BOSCC contexts for both the SC domain, as well as for the Core Total. Inter-rater reliability was lowest for the RRB domain in snack routines (which excluded item 9; Play), but was still adequate (Cicchetti, 1994).

[Insert Table 2. about here]

Comparison across contexts

Pearson correlations and paired-samples *t*-tests were used to evaluate the relationship between child social communication across the play and snack contexts. Consistent with our hypotheses, concurrent ratings of BOSCC-Play and BOSCC-Snack Core Total scores were significantly correlated at all time points (Time 1: $r = 0.50$, $p = 0.001$; Time 2: $r = 0.48$, $p = 0.001$; Time 3: $r = 0.61$, $p < 0.001$). Results were similar for the SC domain (Time 1: $r = 0.48$, $p = 0.001$; Time 2: $r = 0.52$, $p < 0.001$; Time 3: $r = 0.57$, $p < 0.001$). The RRB domain, however, were not consistently correlated across contexts over time (Time 1: $r = 0.24$, $p = 0.12$; Time 2: $r = 0.32$, $p = 0.03$; Time 3: $r = 0.32$, $p = 0.06$). Paired samples *t*-tests were used to examine whether there were mean differences between the BOSCC-Play and BOSCC-Snack domain and Core Total scores at a single time-point (pre-intervention). For this analysis, item 9 (Play) was subtracted from the Core Total and RRB domain, so that scores across the two contexts were comprised of the same items. The SC domain and Core Total did not differ between the two contexts (Table 3), suggesting that the BOSCC captures similar information about child social communication across these two contexts. However, the RRB domain was significantly different, with more RRBs coded in play routines than in snack routines (Table 3).

[Insert Table 3. about here]

Change over time

Multilevel models treating participants as the upper-level unit and trial as the lower-level unit was used to evaluate how the BOSCC changed over time (Pre-intervention, Post-intervention, Follow-up) across the two contexts (Play and Snack). Time and context were treated as fixed effects, and the models included both time and context main effects and their

interaction. Participant variance (i.e., intercept variance) and residual variance were included as random effects. The model was run twice: first, with the all Core Total items for both contexts, and second, with the "Play with Objects" item subtracted such that the same items comprised the Core Totals for both BOSCC-Play and BOSCC-Snack.

Model 1

There was a significant main effect of time, $F(2, 212) = 14.85$, $p < 0.001$. Pairwise comparisons with a Bonferroni correction showed that there was a significant reduction in BOSCC Core Total from Time 1 to Time 2 and from Time 1 to Time 3, but no significant change from Time 2 to Time 3. There was also a significant main effect of context, $F(1, 211) = 42.99$, $p < 0.001$, with BOSCC-Snack scores (M = 25.6, SE = 1.0) being lower than BOSCC-Play scores (M = 30.1, SE = 0.98). The time by context interaction was not significant, $F(2, 210) = 2.69$, $p = 0.07$, suggesting that the BOSCC changed similarly over time in both contexts (Figure 2a).

Model 2

As in the model above, there was a significant main effect of time, $F(2, 214) = 15.06$, $p < 0.001$. Pairwise comparisons with a Bonferroni correction showed that there was a significant reduction in BOSCC Core Total from Time 1 to Time 2 and from Time 1 to Time 3, but no significant change from Time 2 to Time 3. There was also a significant main effect of context, $F(1, 212) = 4.25$, $p = 0.04$, with BOSCC-Snack scores (M = 25.6, SE =0.95) being slightly lower than BOSCC-Play scores (M=27.0, SE =0.93). In contrast with the model above, the time by context interaction was significant, $F(2, 211) = 3.34$, $p = 0.037$. Figure 2b illustrates that while the BOSCC-Play did not appear to change from Time 2 to Time 3, the BOSCC-Snack continued to decrease during this time.

[Insert Figure 2a and 2b. about here]

Correlation with other variables

At Time 1, the BOSCC-Play Core Total was significantly related to the MSEL verbal and nonverbal age equivalents, VABS Communication and Socialization age equivalents, and ADOS Social Affect (SA) and overall algorithm totals. Results for BOSCC-Snack Core Total were similar, with the exception that BOSCC-Snack scores were marginally significantly related to the ADOS SA algorithm total ($p=.05$), but not to the overall algorithm total (Table 4).

[Insert Table 4. about here]

Incremental validity

To evaluate the incremental validity of a second BOSCC context in predicting relevant concurrent outcomes at Time 1, hierarchical multiple regressions were used to evaluate the extent to which BOSCC-Snack Core Total score predicted relevant measures of interest over and above BOSCC-Play Core Total score. In other words, we were interested in whether the addition of a second BOSCC context (i.e. Snack) would contribute significant variance in predicting other commonly-used measures of social communication within a single timepoint.

Based on the significant (and marginally-significant) correlations above, MSEL verbal and nonverbal age equivalents, VABS Communication and Socialization age equivalents, and ADOS SA algorithm total were examined as dependent variables (Table 5). The addition of the BOSCC-Snack Core Total significantly improved prediction of MSEL VAE and NVAE by 8% and 15% variance explained, respectively. Likewise, prediction of the VABS Communication age equivalent was improved by 9% variance explained, which was statistically significant. Addition of the BOSCC-Snack Core Total did not significantly improve prediction of the VABS Socialization age equivalent or the ADOS SA algorithm total.

[Insert Table 5 about here]

Discussion

Overall, results suggest that adapting the BOSCC for use in a home snack routine is feasible and psychometrically sound. Internal consistency and interrater reliability were good, particularly for the SC domain, lending further support for the psychometric properties of the BOSCC. Significant correlations and lack of mean differences between concurrent BOSCC-Play and BOSCC-Snack SC domain scores suggests, in terms of social communication, that the BOSCC captures similar information across caregiver-child interactions in different contexts. This suggests that daily routines such as mealtime provide useful samples of social communication behavior.

However, the properties of the RRB domain were less consistent. Internal consistency and inter-rater reliability were not as strong for the RRB domain. In addition, RRB domain scores across the two contexts were significantly different, with children demonstrating more RRBs during play routines than snack routines. This may be true for several reasons. Children with ASD show more limited play skills compared to typically-developing peers beginning from an early age (Landa, Holman, & Garrett-Mayer, 2007), and may instead participate in sensory exploration or repetitive actions with toys when engaging in an unstructured play interaction (Honey, Leekam, Turner, & McConachie, 2007). In addition, it is possible that RRBs are more observable when they involve an object (e.g. spinning a toy or performing the same play action repeatedly) than when they do not. This is somewhat supported by the pattern of inter-rater reliability found in this study; though ICCs were generally high, the lowest ICCs were for the RRB domain in snack routines, indicating more disagreement among raters in this context. Though the BOSCC captures different amounts of RRBs across two contexts, this alone does not indicate that one context or the other provides an inaccurate estimate of the severity of the child's

RRBs. Rather, this variability may reflect true activity-dependent differences in RRBs, and measurement across contexts may allow for a more accurate representation of functional impairment across settings. Taken together, results suggest that the SC domain has better measurement properties than the RRB domain, consistent with the measure's emphasis on social communication. Additional research focusing on measurement of RRBs across instruments and contexts is needed to clarify the validity of the BOSCC for capturing RRBs.

The BOSCC, in both contexts, was sensitive to change over a relatively brief duration from pre- to post-intervention. This suggests that the BOSCC is useful for evaluating treatment response in intervention trials. At the same time, our results showed that social communication behaviors may change differently across contexts over time. After exclusion of the "Play with Objects" item (which was only scored in play routines), the BOSCC-Snack showed more change than BOSCC-Play, particularly from Time 2 to Time 3. Future research is needed to replicate this result, and to better understand whether the BOSCC is more sensitive to change in social communication skills during particular routines or whether the intervention differentially impacts child skill use across activities.

Measuring child response to intervention in multiple contexts may advance ASD research in several ways. First, response to intervention across contexts is one way to evaluate the generalization of treatment effects for interventions that specifically teach skills in the context of play interactions. Thus, including an observation of caregiver-child interactions during a snack routine using the BOSCC may provide a better estimate of child treatment response to caregiver-mediated interventions.

In addition, this study found that RRBs were less present during snack routines, suggesting that certain behaviors may fluctuate in frequency or severity across contexts, and that

adding the BOSCC-Snack context significantly improved prediction of some concurrent standardized assessment scores. Likewise, we found that change in social communication behaviors differed across contexts. Measuring child skills in multiple contexts may provide a more accurate estimate of the child's functional skills and help capture context-specific changes in social communication. Further, it may allow for a systematic method for evaluating the effect of context on various child behaviors.

One limitation of this study is that we evaluated the utility of the BOSCC in one particular home routine: snack. However, given the similar psychometric properties obtained in play and snack routines it is likely that the BOSCC can be used across a wider range of activities. Future research should extend the use of the BOSCC to other types of daily activities, which may come with different social expectations as well as differences in caregiver behavior. It is also important to understand how caregiver behavior relates to child behavior in these dyadic interactions, given that caregivers likely demonstrate different parenting and communication behaviors in different situations, which in turn, affect child social communication.

References

Anagnostou, E., Jones, N., Huerta, M., Halladay, A. K., Wang, P., Scahill, L., . . . Choi, D. (2015). Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism, 19*(5), 622-636.

Bolte, E. E., & Diehl, J. J. (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.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment, 6*(4), 284.

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. doi:10.1177/1362361315605624

French, L., & Kennedy, E. M. (2018). Annual Research Review: Early intervention for infants and young children with, or at-risk of, autism spectrum disorder: a systematic review. *Journal of Child Psychology and psychiatry*, *59*(4), 444-456.

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). *J Autism Dev Disord, 46*(7), 2464-2479. doi:10.1007/s10803-016-2782-9

Honey, E., Leekam, S., Turner, M., & McConachie, H. (2007). Repetitive behaviour and play in typically developing children and children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(6), 1107-1115.

Ingersoll, B. R., & Dvortcsak, A. (2009). *Teaching Social Communication to Children With Autism: A Practitioner's Guide to Parent Training and a Manual for Parents*: Guilford Press.

Ingersoll, B. R., Wainer, A. L., Berger, N. I., Pickard, K. E., & Bonter, N. (2016). Comparison of a Self-Directed and Therapist-Assisted Telehealth Parent-Mediated Intervention for Children with ASD: A Pilot RCT. *J Autism Dev Disord, 46*(7), 2275-2284. doi:10.1007/s10803-016-2755-z [doi]

10.1007/s10803-016-2755-z [pii]

Jones, R. M., Southerland, A., Hamo, A., Carberry, C., Bridges, C., Nay, S., . . . Rehg, J. M. (2017). Increased eye contact during conversation compared to play in children with autism. *Journal of autism and developmental disorders, 47*(3), 607-614.

Kitzerow, J., Teufel, K., Wilker, C., & Freitag, C. M. (2016). Using the brief observation of social communication change (BOSCC) to measure autism-specific development. *Autism Res, 9*(9), 940-950. doi:10.1002/aur.1588

Landa, R. J., Holman, K. C., & Garrett-Mayer, E. (2007). Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Archives of general psychiatry, 64*(7), 853-864.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism Diagnostic Observation Schedule Second Edition (ADOS-2) Manual (Part 1): Modules 1–4. *Torrance, CA: Western Psychological Services*.

Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., . . . Guthrie, D. (2005). Challenges in evaluating psychosocial interventions for autistic spectrum disorders. *Journal of autism and developmental disorders, 35*(6), 695-708.

McConachie, H., Parr, J. R., Glod, M., Hanratty, J., Livingstone, N., Oono, I. P., . . . Charman, T. (2015). Systematic review of tools to measure outcomes for young children with autism spectrum disorder.

Mullen, E. M. (1995). *Mullen scales of early learning*: AGS Circle Pines, MN.

National Research Council. (2001). *Educating children with autism*: National Academies Press.

Nevill, R. E., Lecavalier, L., & Stratis, E. A. (2016). Meta-analysis of parent-mediated interventions for young children with autism spectrum disorder. *Autism*, 1362361316677838.

Nordahl-Hansen, A., Fletcher-Watson, S., McConachie, H., & Kaale, A. (2016). Relations between specific and global outcome measures in a social-communication intervention for children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 29*, 19-29.

Oono, I. P., Honey, E. J., & McConachie, H. (2013). Parent-mediated early intervention for young children with autism spectrum disorders (ASD). *Cochrane Database Syst Rev*(4), Cd009774. doi:10.1002/14651858.CD009774.pub2

Pijl, M. K., Rommelse, N. N., Hendriks, M., De Korte, M. W., Buitelaar, J. K., & Oosterling, I. J. (2016). Does the Brief Observation of Social Communication Change help moving forward in measuring change in early autism intervention studies? *Autism*, 1362361316669235.

Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., . . . Bruinsma, Y. (2015). Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder. *Journal of autism and developmental disorders, 45*(8), 2411-2428.

Sparrow, S., Cicchetti, D., & Balla, D. (2005). Vineland adaptive behavior scales:(Vineland II), survey interview form/caregiver rating form. *Livonia, MN: Pearson Assessments*.

Stronach, S., & Wetherby, A. M. (2014). Examining restricted and repetitive behaviors in young children with autism spectrum disorder during two observational contexts. *Autism, 18*(2), 127-136.

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2018). Routine Language: Speech Directed to Infants During Home Activities. *Child development*.

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental science, 20*(6), e12456.

Vismara, L. A., & Rogers, S. J. (2010). Behavioral treatments in autism spectrum disorder: what do we know? *Annual review of clinical psychology, 6*, 447-468.

Zwaigenbaum, L., Bauman, M. L., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., . . . Natowicz, M. R. (2015). Early Intervention for Children With Autism Spectrum Disorder Under 3 Years of Age: Recommendations for Practice and Research. *Pediatrics, 136 Suppl 1*, S60-81. doi:10.1542/peds.2014-3667E

Table 1. Participant demographics and descriptive statistics

Child Demographics	n	**Caregiver Demographics**	n
Gender		**Gender**	
Male	37	Male	5
Female	13	Female	45
Race		**Education Level**	
White	39	Some high school	1
Black or African American	4	High school graduate	1
Asian	2	Some college/specialized training	20
More than one race	4	4-year college	18
Other race	1	Graduate degree	10
Ethnicity		**Marital Status**	
Hispanic or Latino	4	Married; living with partner	39
Not Hispanic or Latino	43	Single; divorced or separated	3
Missing	3	Single; living with partner	2
		Single; never married	6

Child Descriptive Statistics	M	SD
Chronological age (months)	46.05	14.50
ADOS CSS	6.85	1.56
MSEL VAE (months)	22.51	11.96
MSEL NVAE (months)	26.70	11.82

Note. ADOS = Autism Diagnostic Observation Schedule; CSS = Calibrated Severity Score; SA = Social Affect, MSEL = Mullen Scales of Early Learning; VAE= Verbal Age Equivalent; NVAE = nonverbal Age Equivalent; VABS = Vineland Adaptive Behavior Scale.

Table 2. Scale descriptives and reliability

	Mean	SD	Skewness	Kurtosis	ICC	α
BOSCC-Play						
Core total	29.85	8.76	-0.09	-0.45	0.88	0.79
SC total	24.45	7.51	-0.14	-0.72	0.87	0.88
RRB total	5.40	3.01	0.81	0.07	0.84	0.46
BOSCC-Snack						
Core total[1]	26.11	8.34	-0.02	-0.50	0.92	0.82
SC total	24.65	7.81	-0.09	-0.67	0.93	0.87
RRB total[1]	1.46	1.80	1.53	2.40	0.70	0.78

Note. [1]RRB & Core items for snack excludes item 9 (Play)

Table 3. Mean differences between BOSCC-Play and BOSCC-Snack at pre-intervention

	BOSCC-Play		BOSCC-Snack			
Domain	M	SD	M	SD	*t*	*p*
Core[1]	28.35	7.99	28.98	7.73	-0.69	0.50
SC	26.06	7.25	27.39	7.16	-1.58	0.12
RRB[1]	2.30	2.24	1.59	2.04	2.24	0.03*

Note: [1] Item 9 (Play) was subtracted from RRB & Core totals for the purpose of this analysis. * $p < 0.05$.

Table 4. Correlations between variables of interest at Time 1.

		ADOS CSS	ADOS Algorithm Total	ADOS SA Algorithm Total	MSEL VAE	MSEL NVAE	VABS Comm. AE	VABS Soc. AE
BOSCC-Play Core total	*r*	0.20	0.41	0.43	-0.67	-0.58	-0.66	-0.49
	p	0.18	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
BOSCC-Snack Core total	*r*	0.05	0.28	0.31	-0.52	-0.58	-0.58	-0.42
	p	0.75	0.08	0.05	<0.001	<0.001	<0.001	0.01
ADOS CSS	*r*	-	0.83	0.75	-0.19	-0.28	-0.09	-0.06
	p		<0.001	<0.001	0.20	0.06	0.53	0.67
ADOS Algorithm Total	*r*		-	0.90	-0.47	-0.49	-0.32	-0.25
	p			<0.001	<0.001	<0.001	0.03	0.09
ADOS SA Algorithm Total	*r*			-	-0.42	-0.42	-0.35	-0.24
	p				<0.001	<0.001	0.02	0.12
MSEL VAE	*r*				-	0.81	0.84	0.60
	p					<0.001	<0.001	<0.001
MSEL NVAE	*r*					-	0.72	0.61
	p						<0.001	<0.001
VABS Comm. AE	*r*						-	0.73
	p							<0.001

Note. * Correlation is significant (*2-tailed* $p < 0.05$). ADOS = Autism Diagnostic Observation Schedule; CSS = Calibrated Severity Score; SA = Social Affect, MSEL = Mullen Scales of Early Learning; VAE= Verbal Age Equivalent; NVAE = nonverbal Age Equivalent; VABS = Vineland Adaptive Behavior Scale.

Table 5. Hierarchical multiple regressions: Concurrent incremental validity of BOSCC-Snack over and above BOSCC-Play at Time 1.

	Model Summary			Coefficients				
	$R^2\Delta$	$F\Delta$	*p*	B	SE	β	*t*	*p*
MSEL VAE								
Model 1 (BOSCC-Play only)	0.41	27.11	<.001*					
Model 2	0.08	5.84	0.021*					
BOSCC-Play				-0.72	0.18	-0.51	-4.03	<.001*
BOSCC-Snack				-0.49	0.20	-0.31	-2.42	0.021*
MSEL NVAE								
Model 1 (BOSCC-Play only)	0.30	16.43	<.001*					
Model 2	0.15	10.35	0.003*					
BOSCC-Play				-0.50	0.18	-0.37	-2.78	0.008*
BOSCC-Snack				-0.66	0.20	-0.43	-3.22	0.003*
VABS Communication AE								
Model 1 (BOSCC-Play only)	0.42	29.13	<0.001*					
Model 2	0.09	7.38	0.010*					
BOSCC-Play				-0.56	0.15	-0.48	-3.79	0.001*
BOSCC-Snack				-0.46	0.17	-0.35	-2.72	0.010*
VABS Socialization AE								
Model 1 (BOSCC-Play only)	0.24	12.27	0.001*					
Model 2	0.05	2.51	0.121					
BOSCC-Play				-0.27	0.11	-0.37	-2.37	0.023*
BOSCC-Snack				-0.20	0.13	-0.25	-1.59	0.121
ADOS SA Algorithm Total								
Model 1 (BOSCC-Play only)	0.19	11.39	0.001*					
Model 2	0.05	2.83	0.099					
BOSCC-Play				0.13	0.06	0.32	2.16	0.036*
BOSCC-Snack				0.11	0.07	0.25	1.68	0.099

Note. * $p < 0.05$; MSEL = Mullen Scales of Early Learning; VAE = verbal age equivalent; NVAE = non-verbal age equivalent; VABS = Vineland Adaptive Behavior Scales; AE = age equivalent.

Figure Legends

Figure 1. Distribution of BOSCC-Snack and BOSCC-Play Core Total and domain scores.

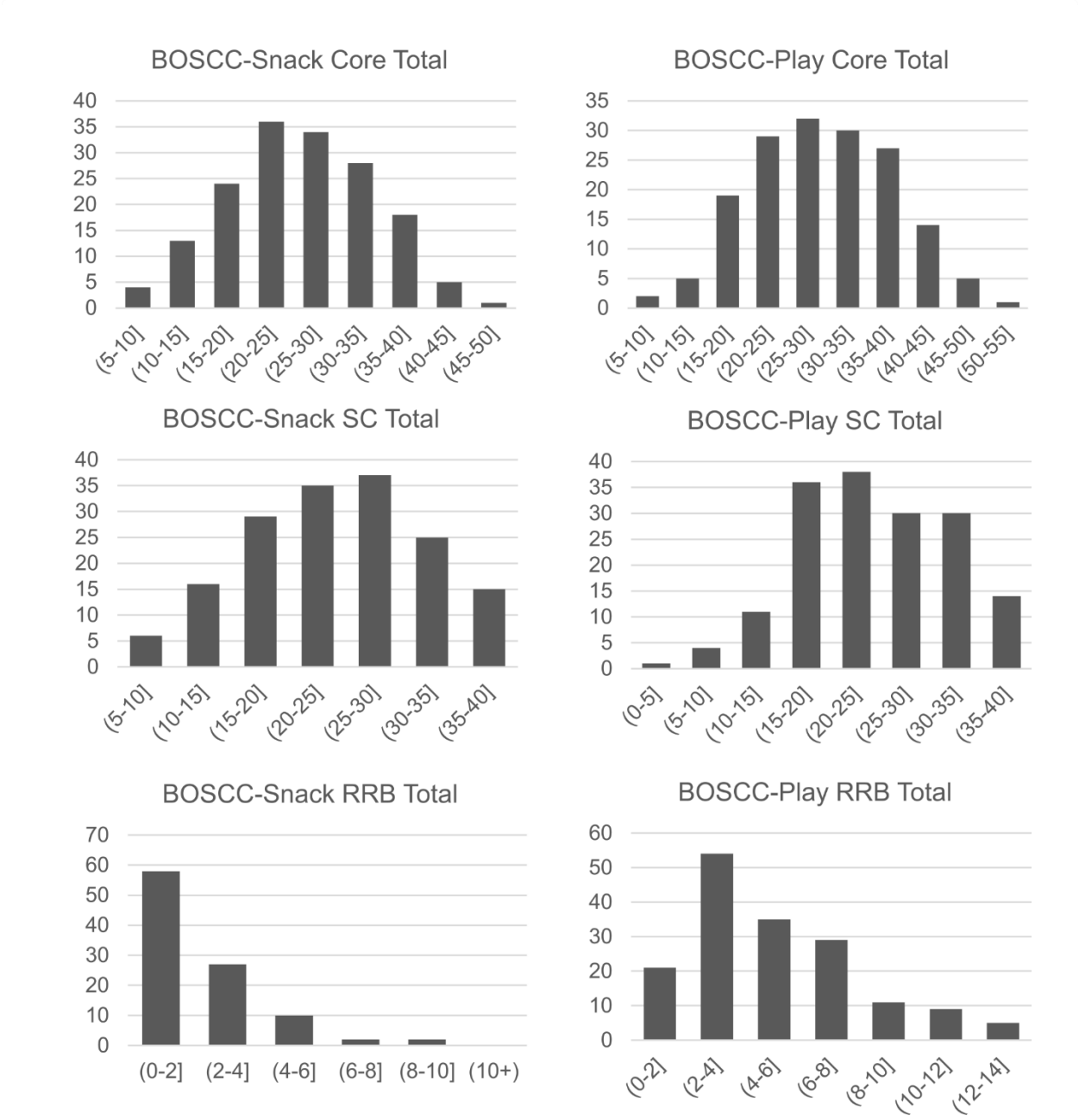

Figure 1. Distribution of BOSCC-Snack and BOSCC-Play Core Total and domain scores.

Figure 2a. Estimated marginal means of BOSCC-Play and BOSCC-Snack Core Totals over time. Error bars represent 95% confidence intervals.

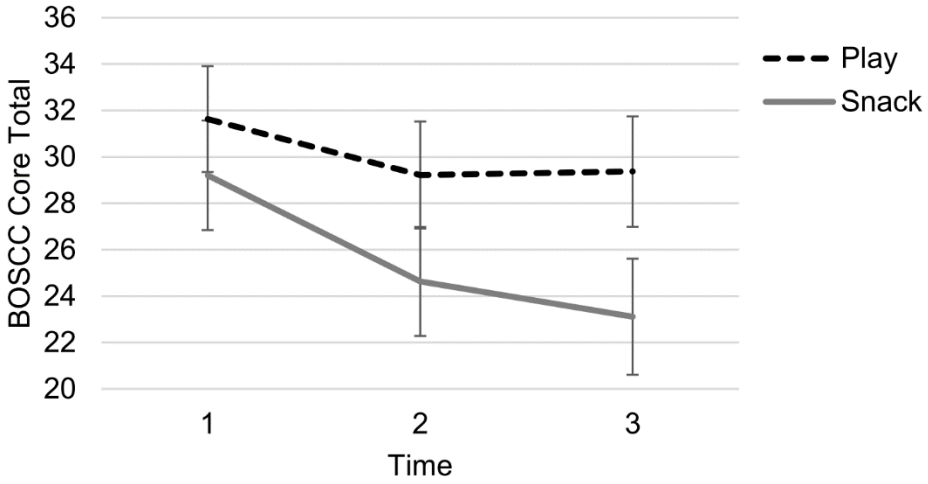

Figure 2a. Estimated marginal means of BOSCC-Play and BOSCC-Snack Core Totals over time. Error bars represent 95% confidence intervals.

Figure 2b. Estimated marginal means of BOSCC-Play (with Play item subtracted) and BOSCC-Snack Core Totals over time. Error bars represent 95% confidence intervals.

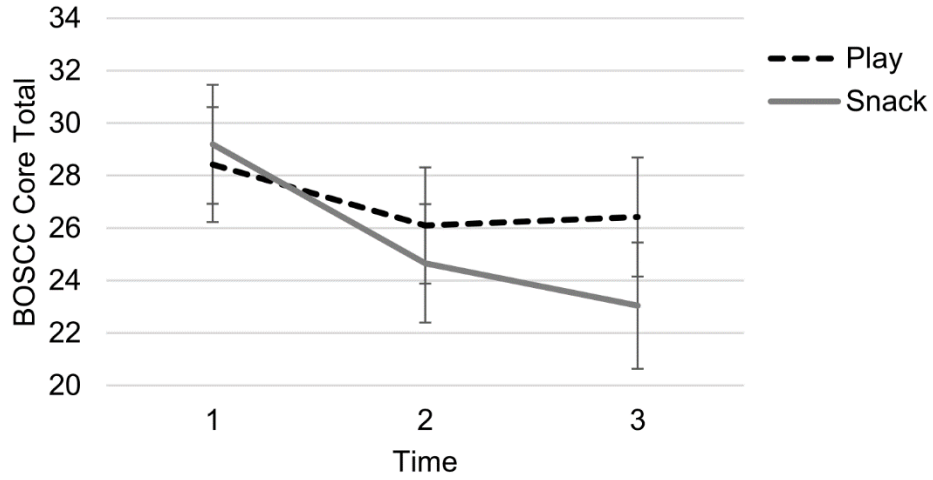

Figure 2b. Estimated marginal means of BOSCC-Play (with Play item subtracted) and BOSCC-Snack Core Totals over time. Error bars represent 95% confidence intervals.