Using thin-slice ratings to measure social communication in children with autism spectrum disorder.

Kyle M. Frost[1], Kaylin M. Russell[1], & Brooke Ingersoll[1]

[1] Department of Psychology, Michigan State University, East Lansing, MI, USA

Corresponding author:

Brooke Ingersoll

Department of Psychology, Michigan State University, East Lansing, MI, USA

ingers19@msu.edu

(517) 432-8412

This is a post-print version of [Frost, K. M., Russell, K. M., & Ingersoll, B. (2020). Using thin slice ratings to measure social communication in children with autism spectrum disorder. *Research in Autism Spectrum Disorders*.]; the publisher's version can be located at https://doi.org/10.1016/j.rasd.2020.101550.

Abstract

Background: Thin-slice ratings have been shown to predict a variety of behaviors in children and adults. Recently, thin-slice ratings have been used to characterize social communication and autism symptomatology in children with autism spectrum disorder (ASD), suggesting their potential as an alternative to intensive observational coding systems.

Objectives: This study examined whether thin-slice ratings of social communication skills during a parent-child interaction are psychometrically valid, related to observational and standardized measures of child social communication, and sensitive to change over time in children with ASD.

Methods: Thin-slice ratings for 71 children were completed by 173 undergraduates from a psychology research pool. Groups of naïve raters viewed two-minute clips of parent-child interactions for different children at two time points. For each clip, raters assigned scores for 7 items derived from the Brief Observation of Social Communication Change (BOSCC).

Results: Thin-slice rating items formed a unidimensional scale with good internal consistency (α=.92) and inter-rater reliability. The thin-slice rating scale demonstrated convergence with observational data from the BOSCC, as well as other common measures of social communication. Unlike the BOSCC, the thin-slice ratings did not change significantly from Time 1 to Time 2.

Conclusion: Thin-slice ratings of young children with ASD during a parent-child interaction demonstrated criterion validity with the BOSCC and convergent validity with other measures of child social communication functioning. Future studies should investigate whether thin-slice ratings by naïve raters capture other aspects of autism symptomatology. Findings suggest thin-slice ratings may provide a stable estimate of child social communication functioning that tracks with other measures of child developmental functioning.

Keywords: ASD, social communication, thin-slice ratings, observational measures.

Using thin slice ratings to measure social communication in children with autism spectrum disorder.

Children with autism spectrum disorder (ASD) experience deficits in social interaction, which are linked to difficulty with verbal and nonverbal communication (Grossman, 2015). Behavioral observations during natural and semi-structured interactions with a social partner provide an important source of information regarding social communication behaviors in young children with ASD. Observational rating schemes used to quantify these behaviors in a valid and reliable manner often require extensive training, expertise, and time to implement. Thus, while the information obtained is clinically valuable, the resources required can make this type of data collection prohibitive.

Thin slice ratings (TSR; Ambady & Rosenthal, 1992), in which multiple untrained raters provide "gist" ratings of brief segments of expressive behavior (i.e., usually less than 5 minutes), may offer an alternative to more intensive coding measures. Thin-slice ratings have been shown to predict a variety of personality traits, internal states, and social interactions in both children and adults (see Ambady et al., 2000 for a review). Studies have found thin-slice ratings to be representative of behavior across a longer behavioral stream (Murphy, 2005; Prime et al., 2014), and concordant with expert ratings using more intensive behavioral coding systems (Baker et al., 2010; Prime et al., 2014). Thus, the use of thin-slice ratings has the potential to save time, energy, and money when compared to more intensive coding approaches (Baker et al., 2010; Prime et al., 2014).

Recently, research has begun to examine the potential of using thin-slice ratings to quantify social communication and autism symptomatology in young children with ASD. Hampton et al. (2019) found that thin-slice ratings of autism symptoms taken from a language sample and parent-child interaction were significantly correlated with standardized assessments of language and autism severity in a sample of 60 toddlers with and without ASD. In addition, thin-slice ratings were effective in distinguishing children with ASD from children with developmental language disorder and typical development, indicating that thin-slice ratings may provide important information regarding social communication development and autism risk. Walton and Ingersoll (2016) found that thin-slice ratings of child social communication skills in 22 preschool-aged children with ASD were correlated with scores on a standardized assessment of language as well as measures of imitation and joint attention. These studies suggest that thin-slice ratings

show potential as an alternative method for measuring social communication skills in young children with ASD. However, replication of the relationship between thin-slice ratings and standardized measures of social communication in larger samples is needed. Further, research has not yet demonstrated how thin-slice ratings might compare to other observational rating schemes in ASD, or whether these ratings are sensitive enough to capture change over time.

The Brief Observation of Communication Change (BOSCC; Grzadzinski et al., 2016) is standardized 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. Validation of the BOSCC suggests that it is reliable across raters and brief time intervals and shows convergent validity with other measures of social communication (Frost et al., 2019; Grzadzinski et al., 2016; Kim et al., 2019; Kitzerow et al., 2016). Importantly, the BOSCC is more sensitive than the ADOS for capturing change over time (Frost et al., 2019; Grzadzinski et al., 2016; Kim et al., 2019). However, like many observational rating schemes, the training and coding processes for the BOSCC are highly resource intensive (Frost et al., 2019; Grzadzinski et al., 2016).

The current study had three aims relating to understanding the utility of thin slice ratings for measuring social communication in research. First, we wanted to capture the criterion validity of thin-slice ratings by comparing it to a validated observational rating scheme, the BOSCC. Next, we wanted to replicate previous studies demonstrating the relationship between thin-slice ratings and standardized measures of social communication in young children with ASD using a larger sample. Last, we wanted to identify whether thin-slice ratings were sensitive to short-term change in social communication. We predicted that thin-slice ratings would be psychometrically valid and significantly related to other standardized measures of social communication, including the BOSCC. We also expected that the thin-slice ratings would be sensitive to change in social communication over a 4-6 month interval, as evidenced by significant differences between two time points.

Method

Procedures

Groups of naïve raters viewed two-minute audio-video clips of parent-child free play interactions filmed in families' homes at intake (T1) and 4-6 months later (T2). Parents were provided with standardized box of toys and were told to play with their child for ten-minutes as they normally do, and the interaction was filmed. The two-minute thin slice clips were selected from minutes 4:00 through 6:00 of original 10-minute videos, in line with previous studies (e.g. Walton & Ingersoll, 2016), in order to obtain a sample that brief yet sufficient to contain several examples of child social communication. Fourteen groups of 11-15 raters viewed 8-11 video clips each, such that no group saw two videos of the same child. This number of raters was selected based on a review of the literature on thin-slice ratings. Literature in ASD has used a widely variable number of raters (e.g. Hampton et al., 2019; Sasson et al., 2017; Walton & Ingersoll, 2016), and there is no clear consensus on the optimal number. While Hampton and colleagues found that 5 raters could achieve sufficient stability (2019), their raters had a relatively high level of education and coding experience and they rated multiple contexts, therefore we opted to include a larger number of raters. Raters were informed that they would view several brief clips of young children and their parents playing together. Before viewing the clips, raters were told to keep in mind that they would be providing their opinions on seven questions after viewing each clip. They were not given any additional training or instruction. The administrator did not describe the explicit purpose of the study until the debriefing period.

Participants

Video Raters. Thin-slice ratings were completed by 173 undergraduate college students from a psychology research subject pool. Most raters were women (78%) and averaged 19.2 years old (SD = 1.12). In terms of racial identity, 2.3% of participants identified as more than one race. Of the remaining, 69.2% identified as White, 13.6% identified as Black, 12.4% identified as Asian, 1.2% identified as American Indian or Alaska Native, and 3.6% selected "Other." In terms of ethnicity, 5.8% participants identified as Hispanic/Latinx. Most students were in their first (34.1%) or second (32.9%) year of college. Most were majors in social sciences (58.4%), followed by medical and health sciences (20.2%) and natural sciences (11.6%).

Raters varied in their self-reported familiarity with ASD; 25% reported having an immediate or extended family member with ASD, and 16% reported having a friend with ASD. Thirty-nine percent said

they had learned about ASD in the context of an internship, job, volunteer position, or professional workshop, and 50% said they had learned about it in a class. About a quarter (28%) reported no prior knowledge of or experience with ASD.

Video Subjects. Participants in this study were 71 children (76% male) with ASD between the ages of 17 and 93 months (M = 46.11, SD = 15.97) and their caregivers. In terms of racial identity, children were 73% White, 13% multi-racial, 7% Black or African-American, 6% Asian or Asian-American, and 1% another race. Children had early social communication delays, as evidenced by parent-reported communication age-equivalent scores (M = 26.55 months, SD = 11.38) and expressive language age equivalent on a standardized cognitive test (M = 21.06 months, SD = 10.92). Families in Michigan participated in one of two research studies with similar inclusion criteria evaluating the efficacy of a telehealth-based parent-mediated intervention (Ingersoll et al., 2016). All children had a classification of "autism" or "autism spectrum" on the Autism Diagnostic Observation Schedule-Second Edition (ADOS; Lord et al., 2012) and had an expressive language age equivalent at or below 48 months on the Mullen Scales of Early Learning (MSEL; Mullen, 1995) . The only difference in inclusion criteria was that the second study expanded the eligible age range from 6 years and younger to 8 years and younger. Twenty-six participants were enrolled in a pilot randomized-controlled trial (RCT) from 2012 to 2014, and 45 participants were enrolled in an ongoing full-scale RCT from 2015 to 2018. All participants, regardless of treatment allocation, were included in this study, as we expected all children to demonstrate some developmental change over the course of 4-6 months.

Measures

Thin-Slice Ratings. The 7 thin-slice rating items were adapted from the social communication scale of the BOSCC (see Table 1). The BOSCC item pertaining to integration of verbal and nonverbal communication was not included in the thin-slice rating items due to the complex focus of the item. Raters used a 5-point Likert scale ranging from (1) strongly disagree to (5) strongly agree to rate all items for each clip, with higher scores indicating less impairment. Each child's Thin-Slice Rating Item Scores were calculated by averaging across all available ratings for each item. Each child's Thin-Slice Rating Overall Score was calculated by averaging across Thin-Slice Rating Item Scores.

BOSCC. A pre-publication research version of the BOSCC (Version 8/17/16; Grzadzinski et al., 2016) was used to rate the original 10-minute parent-child free play interactions at T1 and T2 with permission from WPS and measure authors. This study utilized the first 8 items comprising the social communication (SC) domain. Each item is rated on a six-point scale using a detailed decision tree, with higher scores indicating greater impairment. Primary raters were research assistants who were research reliably trained to code BOSCC. These research assistants were blind to child treatment status and video time point. Raters sometimes coded multiple videos of the same child, but did not code videos of the same child consecutively. The videos used for this paper comprise a subset of those on which we previously published BOSCC validity data (Frost et al., 2019). Sixty percent of videos in this sample were double-coded; a single-measures, two-way mixed design using absolute agreement yielded an intraclass correlation of .78.

The MacArthur-Bates Communicative Development Inventory (MCDI). The MCDI is a parent-report checklist used to determine words that children understand and say (Fenson, 2007) at T1. Parents were asked to complete either the Words and Gestures version or the Words and Sentences form depending on their child's language ability. The number of words produced was used to measure expressive language.

Vineland Adaptive Behavior Scales, 2nd Edition (VABS). The VABS is a semi-structured parent interview designed to evaluate adaptive functioning (Sparrow et al., 2005), with higher scores indicating less impairment. The VABS was conducted by phone with parents to assess children's adaptive skills at T1; the Communication and Socialization domain raw sums were used for this study.

Mullen Scales of Early Learning (MSEL). The MSEL is a standardized cognitive assessment which was administered at T1 to evaluate the developmental level of participants (Mullen, 1995). The MSEL provides scores across four subscales. A verbal raw sum was calculated for each child by adding the Receptive Language and Expressive Language subscales raw scores. A nonverbal raw sum was calculated for each child by adding the Fine Motor and Visual Reception subscales raw scores.

ADOS-2. The ADOS-2, a play-based clinical assessment of core autism symptomatology that evaluates social affect and the presence of restricted and repetitive behaviors (Lord et al., 2012), was administered at T1. Modules 1, 2, or Toddler were administered to participants for this study based on the

child's expressive language level. The ADOS Calibrated Severity Score (CSS) provides an estimate of autism severity ranging from 1-10 with higher scores indicating more prominent autism symptoms (Gotham et al., 2009).

Analysis Plan

An exploratory factor analysis was used to determine the factor structure of the scale, and Cronbach's alpha was used to examine internal consistency. Intraclass correlations (ICCs) were examined to evaluate inter-rater reliability. Bivariate and partial correlations were run to examine the relationship between Thin-Slice Rating Overall Score and other measures of social communication development, including standardized measures and BOSCC ratings for the longer parent-child free play interaction. To assess divergent validity, we examined the relationships between Thin-Slice Rating Overall Score and child sex using a point biserial correlation, and Thin-Slice Rating Overall Score and parent education level using a Spearman rank-order correlation. Last, paired-samples t-tests were used to determine whether the Thin-Slice Rating Overall Score captured short-term change similarly to the BOSCC, which has been demonstrated to change in previous studies (Frost et al., 2019; Grzadzinski et al., 2016).

Results

Scale reliability

An exploratory factor analysis with principal axis factoring supports the conclusion that these items form a unidimensional scale. Examination of the scree plot and the magnitude of eigenvalues were used to determine the best factor solution. All sources of information indicated that a one-factor solution provided the best fit to the data, and thus a single factor with an eigenvalue of 4.83 was extracted. Factor loadings ranged from .69 to .90. Cronbach's alpha for the 7 Thin-Slice Rating Item Scores was .92, indicating high internal consistency. Inter-item correlations ranged from .46 to .92. Separate one-way random average measures ICCs were run for the Thin-Slice Rating Overall Score for each group of raters because groups had different numbers of randomly selected raters. ICCs were used because they provide an estimate of agreement between raters that accounts for both consistency and absolute value. ICCs indicated excellent reliability, ranging from .77 to .98, with an average ICC of .91 (Table 2; Cicchetti, 1994).

Criterion validity

The Thin-Slice Rating Overall Score at T1 was significantly correlated with the T1 BOSCC Social Communication total with a large effect size (r = -.69, p < .001; Table 3). Both thin-slice ratings and BOSCC were significantly correlated with chronological age (r = .27, p = .021 and r = -.31, p = .014, respectively) and MSEL nonverbal raw sum (r = .41, p = .001 and r = -.46, p < .001, respectively). After controlling for age and MSEL nonverbal raw sum, the relationship between the Thin-Slice Rating Overall Score and BOSCC remained significant (r = -.62, p < .001) with a large effect size (Table 3).

Convergent and divergent validity

The Thin-Slice Rating Overall score and BOSCC at T1 were both significantly correlated with parent-reported number of words produced from the MCDI (r = .47, p < .001 and r = -.57, p < .001, respectively), the MSEL verbal raw sum (r = .46, p < .001 and r = -.58, p < .001, respectively), and the VABS Communication raw sum (r = .40, p = .001 and r = -.58, p < .001, respectively). Neither were significantly correlated with the ADOS CSS (r = -.11, p = .366 and r = .04, p = .745 respectively) or the Vineland Socialization raw sum (r = .19, p = .127 and r = -.21, p = .119 respectively). After controlling for age and NVIQ, most of these relationships reduced in magnitude but remained significant; however, the relationships between thin-slice ratings and MCDI (r = .21, p = .160) and MSEL (r = .22, p = .148) were no longer significant (Table 3). In terms of divergent validity, the Thin-Slice Rating Overall Score was not significantly correlated with child sex (r_{bis} = .001, p = .994) or parent education level (r_s = .15, p = .207).

Sensitivity to change

A paired samples t-test showed no significant change in Thin-Slice Rating Overall Score from T1 (M = 3.21, SD = .63) to T2 (M = 3.22, SD = .72), t(58)= -.25, p =.804, Cohen's d =.013, in contrast to the BOSCC Social Communication total (T1: M = 26.71, SD = 6.90; T2: M = 23.70, SD = 6.46), t(58) = 3.85, p < .001, Cohen's d =.439.

Discussion

The goal of this study was to examine the utility of thin-slice ratings methodology for measuring child social communication skills during parent-child interactions. Results show that the thin-slice rating items we examined formed a unidimensional scale with high internal consistency and average-measures inter-rater reliability. The Thin-Slice Rating Overall Score was related in the expected direction to the

BOSCC, which is a much more intensive video rating scheme, suggesting that similar information can be captured in a shorter time. We also replicated the finding that thin-slice ratings related to broader measures of developmental level in the expected direction. Even when controlling for both age and NVIQ, the Thin-Slice Rating Overall Score was significantly related to less similar measures of child social communication than the BOSCC, including a standardized parent-reported interview of adaptive communication and socialization skills. This relationship suggests that thin-slice ratings by naïve raters capture meaningful variation in child social communication ability, replicating previous findings. However, unlike the BOSCC, the Thin-Slice Rating Overall Score did not capture short-term change in child social communication. This suggests that, although some change in social communication occurred, the thin-slice rating does not provide a precise enough estimate of skills to capture change.

Thin-slice rating has been proposed as a useful methodology due to the ease with which large numbers of ratings can be obtained. They do not require many hours of training as do more complex observational rating schemes, or the vast number of hours required for moment-to-moment coding. Our study demonstrates that thin-slice ratings capture meaningful variation in child social communication skills that tracks with similar yet more intensive coding schemes as well as standardized measures of social communication. However, the utility of thin-slice ratings as implemented in this study is limited insofar as these ratings are not precise enough to capture change. Yet, it is possible that some changes to the thin slice methods would improve sensitivity to change. For example, a longer observation, multiple samples across different contexts or providing limited training of the naïve raters could provide greater reliability and therefore more sensitivity to change.

This study has several limitations which affect the interpretability of results. Our naïve raters were sampled from a psychology undergraduate subject pool, and thus results may not generalize to naïve raters drawn from other populations. A more diverse sample, in terms of age, gender, education level, and racial-ethnic background may yield different results. In addition, we did not collect information on whether raters had any prior coding experience or experience rating social communication which may affect their responses.

Implications

This research shows that ratings of child social communication by naïve observers track with other established measures but do not capture short-term improvement. While these data show that thin-slice ratings capture meaningful information, the procedure may need to be adapted in order to improve sensitivity to change. Future research may attempt to provide limited training to raters in thin slice studies, show raters multiple videos of the same child, or use longer behavioral observations to see if the precision of subtle social behaviors increases sufficiently to capture subtle changes in social communication skills. This would enable the efficient observational coding of subtle communication behaviors.

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 the funding agencies. We thank the undergraduate research assistants who assisted with data collection and the families who participated in this study.

Conflict of Interest Statement

The last author receives royalties from the sale of the manual used in the research. Royalties are donated to the research. There are no conflicts of interest to report.

Table 1. Thin-slice ratings and corresponding BOSCC Social Communication items.

TSR Items	**Corresponding BOSCC Items**
1) The child looks at the adult in order to communicate	Eye Contact
2) The child shows facial expressions	Facial Expressions
3) The child use gestures to communicate	Gestures
4) The child vocalizes to the adult to communicate	Vocalizations Directed to Others
5) The child initiates or directs the adult's attention in any manner	Frequency and Function of Social Overtures
6) The child responds to the adult's attempts to engage with the child	Frequency and Quality of Responses
7) The child engages in interactive play with the adult (e.g., turn-taking, imitating, social routines), beyond just observing the adult's actions.	Engagement in Play activities/Interaction

Note. TSR = Thin Slice Ratings, BOSCC = Brief Observation of Social Communication Change.

Table 2. Inter-rater reliability across groups of naïve raters, including the number of raters per group and the number of videos on which ratings were completed.

Group	n Raters	n Videos observed by group	Average measures intraclass correlation
A	11	10	0.85
B	12	10	0.84
C	15	9	0.92
D	11	9	0.94
E	12	10	0.96
F	12	10	0.96
G	12	11	0.77
H	12	10	0.93
I	14	10	0.98
J	13	8	0.91
K	11	8	0.93
L	11	8	0.96
M	13	8	0.90
N	14	8	0.97

Table 3. Bivariate correlations and partial correlations controlling for age and MSEL nonverbal raw sum between thin-slice ratings and other measures at Time 1.

		2	3	4	5	6	7
1. TSR Overall Score	bivariate	-.69*	.47*	-.11	.46*	.40*	0.19
	partial	-.62*	.20	-.09	.20	.24	.30*
2. BOSCC SC Total	bivariate		-.57*	.04	-.58*	-.54*	-.21
	partial		-.37*	-.07	-.42*	-.41*	-.29*
3. MCDI Words Produced	bivariate			.04	.89*	.63*	.22
	partial			.14	.83*	.55*	.51*
4. ADOS CSS	bivariate				-.02	-.04	-.02
	partial				.14	.10	.08
5. MSEL verbal raw sum	bivariate					.67*	.18
	partial					.59*	.46*
6. VABS Communication raw sum	bivariate						.22
	partial						.47*
7. VABS Socialization raw sum							

Note. TSR = Thin slice rating; BOSCC = Brief Observation of Social Communication Change; SC = social communication; ADOS = autism diagnostic observation schedule; CSS = calibrated severity score; MCDI = MacArthur-Bates Communicative Development Inventory; MSEL = Mullen Scales of Early Learning; VABS = Vineland Adaptive Behavior Scales; * = $p < .05$.

References

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In *Advances in experimental social psychology* (Vol. 32, pp. 201–271). Elsevier.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*(2), 256.

Baker, J. K., Haltigan, J. D., Brewster, R., Jaccard, J., & Messinger, D. (2010). Non-expert ratings of infant and parent emotion: Concordance with expert coding and relevance to early autism risk. *International Journal of Behavioral Development*, *34*(1), 88–95. https://doi.org/10.1177/0165025409350365

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

Fenson, L. (2007). *MacArthur-Bates communicative development inventories*. Paul H. Brookes Publishing Company Baltimore, MD.

Frost, K. M., Koehn, G. N., Russell, K. M., & Ingersoll, B. (2019). Measuring child social communication across contexts: Similarities and differences across play and snack routines: Measuring social communication across contexts. *Autism Research*. https://doi.org/10.1002/aur.2077

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.

Grossman, R. B. (2015). Judgments of social awkwardness from brief exposure to children with and without high-functioning autism. *Autism*, *19*(5), 580–587. https://doi.org/10.1177/1362361314536937

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. https://doi.org/10.1007/s10803-016-2782-9

Hampton, L. H., Curtis, P. R., & Roberts, M. Y. (2019). *Autism at a glance: A pilot study optimizing thin-slice observations*. *23*(4), 971–979.

Ingersoll, B., 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. *Journal of Autism and Developmental Disorders*, *46*(7), 2275–2284. https://doi.org/10.1007/s10803-016-2755-z

Kim, S. H., Grzadzinski, R., Martinez, K., & Lord, C. (2019). Measuring treatment response in children with autism spectrum disorder: Applications of the Brief Observation of Social Communication Change to the Autism Diagnostic Observation Schedule. *Autism*, *23*(5), 1176–1185. https://doi.org/10.1177/1362361318793253

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

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule–2nd edition (ADOS-2). *Los Angeles, CA: Western Psychological Corporation*.

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

Murphy, N. A. (2005). Using Thin Slices for Behavioral Coding. *Journal of Nonverbal Behavior*, *29*(4), 235–246. https://doi.org/10.1007/s10919-005-7722-x

Prime, H., Perlman, M., Tackett, J. L., & Jenkins, J. M. (2014). Cognitive Sensitivity in Sibling Interactions: Development of the Construct and Comparison of Two Coding Methodologies. *Early Education and Development*, *25*(2), 240–258. https://doi.org/10.1080/10409289.2013.821313

Sasson, N. J., Faso, D. J., Nugent, J., Lovell, S., Kennedy, D. P., & Grossman, R. B. (2017). Neurotypical Peers are Less Willing to Interact with Those with Autism based on Thin Slice Judgments. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep40700

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

Walton, K. M., & Ingersoll, B. (2016). The utility of Thin Slice ratings for predicting language growth in children with autism spectrum disorder. *Autism*, *20*(3), 374–380.