Brief Report: Calculation and Convergent and Divergent Validity of a New ADOS-2 Expressive Language Score

Micah O. Mazurek

Mary Baker-Ericzén

Stephen M. Kanne

Corresponding Author:

Micah O. Mazurek, PhD

E-mail: mazurekm@virginia.edu

Department of Human Services

University of Virginia

417 Emmet Street South

Charlottesville, VA 22903, USA

Mary Baker-Ericzén, PhD

Child and Adolescent Services Research Center

Autism Discovery Institute

Rady Children's Hospital

San Diego, CA, USA

Stephen M. Kanne, PhD

Department of Health Psychology

Thompson Center for Autism and Neurodevelopmental Disorders

University of Missouri

Columbia, MO, USA

Acknowledgements

The authors are extremely grateful to all the families who participated in this study and to the members of the study team. Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under award number R01MH097726. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Brief Report: Calculation and Convergent and Divergent Validity of a New ADOS-2 Expressive Language Score

Running Title: ADOS-2 Expressive Language Score

Abstract

Despite the importance of expressive language for individuals with autism spectrum disorder (ASD), expressive language assessments are not consistently included in ASD research and many studies do not adequately describe participants' verbal abilities. A valid and efficient measure of expressive language would facilitate consistent reporting across ASD research studies and provide data for additional analyses. The current study developed a new ADOS-2 Expressive Language Score and examined convergent and divergent validity in a large, well-defined sample of children with ASD. This score was highly correlated with other measures of expressive language (including parent-report, direct assessment, and clinician ratings) and less strongly correlated with measures of receptive language and nonverbal cognitive ability, providing good evidence of convergent and divergent validity.

Keywords

Autism spectrum disorder; autism; communication; verbal ability; assessment; measurement

Impairment in social communication is a core feature of autism spectrum disorder (ASD) (American Psychiatric Association, 2013), and language delays and deficits are common among children with ASD (Kwok, Brown, Smyth, & Cardy, 2015; Weismer, Lord, & Esler, 2010; Wetherby et al., 2004). A substantial proportion of individuals with ASD (30-40%) remain minimally verbal into adulthood (Howlin, Savage, Moss, Tempier, & Rutter, 2014; Pickles, Anderson, & Lord, 2014); however, there is considerable variability in language functioning across the ASD population (Rapin, Dunn, Allen, Stevens, & Fein, 2009; Tager-Flusberg, 2006). As a result, DSM-5 diagnostic criteria now include a specifier indicating whether or not accompanying language impairment is present (American Psychiatric Association, 2013). Language is important for daily functioning across contexts, and early language development is predictive of a variety of outcomes among individuals with ASD (Howlin, 2003; Mayo, Chlebowski, Fein, & Eigsti, 2013). Thus, verbal ability is critically important to assess in terms of both clinical relevance and phenotypic characterization.

Despite the importance of characterizing verbal ability among individuals with ASD, this is not currently standard practice across the majority of ASD research studies. While many studies include diagnostic assessments and IQ scores (or general categories such as “high” or “low” functioning), measures of language are not uniformly administered or reported. Even basic descriptions of verbal ability are often lacking in sample descriptions. When language is assessed, methods and measures vary widely across studies, making it difficult to compare results. This greatly limits our understanding of whether study samples are representative of the broader ASD population, the extent to which findings are generalizable, and the extent to which language may relate to key study variables or outcomes.

One reason for inconsistent reporting practices may be a general lack of consensus regarding how to assess and define language and verbal ability in ASD research (Bal, Katz, Bishop, & Krasileva, 2016; Plesa Skwerer, Jordan, Brukilacchio, & Tager-Flusberg, 2016). When considering appropriate measures of language and verbal ability, currently available options have both advantages and disadvantages. Standardized language assessments are norm-referenced, provide the ability to assess language across a range of domains (from receptive to expressive), and allow comparisons of performance to that of age-matched peers. The Clinical Evaluation of Language Fundamentals (CELF) (Wiig, Semel, & Secord, 2003; Wiig, Secord, & Semel, 2004), the Comprehensive Assessment of Spoken Language (CASL) (Carrow-Woolfolk, 2017) and the Preschool Language Scales (PLS) (Zimmerman, Steiner, & Pond, 2011) are examples of widely-used standardized language measures. However, standardized testing of language in children with ASD is associated with a number of challenges. For example, the performance of children with very low language ability may fall below basal levels on many standardized tests (Charman, 2004). Autism-specific characteristics, such as limited social engagement, lack of understanding of task demands, repetitive behaviors, and echolalia may also interfere with performance on formal language assessments (Koegel, Koegel, & Smith, 1997; Tager-Flusberg, 2000). Parent-report measures of language ability, such as the Children's Communication Checklist 2nd Edition (Bishop, 2003) or the MacArthur Communicative Development Inventory (CDI) (Fenson et al., 1993; Fenson et al., 1994), address many of these concerns, but may not be appropriate for individuals across ages and language levels and may be subject to reporter bias (Stiles, 1994).

In contrast to standardized measures, natural language samples provide assessments of language use in informal settings and unstructured contexts. For example, some studies of

expressive language in children with ASD have used language sampling within the context of unstructured play-based interactions with either parents or examiners (Eigsti, Bennetto, & Dadlani, 2007; Hale & Tager-Flusberg, 2005; Swensen, Kelley, Fein, & Naigles, 2007; Tager-Flusberg et al., 1990). In studies of typically developing children, measures of spontaneous speech derived from natural language samples have been found to correlate significantly with scores on standardized assessments (Bornstein & Haynes, 1998; Ukrainetz & Blomquist, 2002). Similar results have been found in research on children with ASD, including significant correlations between spontaneous speech assessed through natural language samples and scores on the Clinical Evaluation of Language Fundamentals (CELF), Peabody Picture Vocabulary Test-Third Edition (PPVT-III), and Expressive Vocabulary Test (Condouris, Meyer, & Tager-Flusberg, 2003). However, natural language sampling requires both technical expertise and substantial time and resources devoted to recording, transcription, and coding. Natural language sampling is generally conducted by speech language pathologists or other professionals with expertise in language and communication, although it may also be conducted by graduate students or other trained examiners. It also requires an adequate and representative sample of language. In general, the sample should be at least 30 minutes in length and/or 50 utterances (Cole, Mills, & Dale, 1989; Lund & Duchan, 1993; Tager-Flusberg, et al., 2009; Thomas, 1989); however, there is some evidence that briefer language samples can be useful in some cases (Heilmann, Nockerts, & Miller, 2010). Natural language sampling also requires the ability to record and transcribe the session(s), and the technical skill to accurately and reliably code the child's utterances (Cole, Mills, & Dale, 1989; Lund & Duchan, 1993; Tager-Flusberg, et al., 2009; Thomas, 1989). As such, it is often not feasible in many clinical and research contexts.

Alternative options for valid and efficient assessment of expressive language functioning would be useful for fostering consistency in reporting across ASD research studies. Given the abovementioned barriers to incorporating comprehensive language assessments into typical research protocols, it may be worthwhile to consider feasible options that could be easily incorporated into standard ASD assessment batteries.

Among assessment tools reported in current ASD research, the Autism Diagnostic Observation Schedule – 2nd Edition (ADOS-2) (Lord et al., 2012) is arguably one of the most widely used ASD diagnostic tools. The ADOS-2 is frequently included in ASD research for diagnostic confirmation and reporting of sample characteristics. It is included in the list of NIH-established common data elements for ASD research and is a required measure for the trans-NIH Autism Centers of Excellence (ACE) program, which funds large-scale multidisciplinary ASD research conducted through ASD research centers and networks. The ADOS-2 has the additional benefit of being appropriate for individuals across ages and language levels as it has separate modules available to assess toddlers through adults. All modules of the ADOS-2 contain scored items that focus on language and communication, highlighting its potential utility for generating a common language metric.

Unfortunately, prior studies have not fully utilized the ADOS item scores as a measure of language due to the lack of an ADOS-derived calculated standardized score. A few researchers, however, used language samples gathered during ADOS administration. For example, Park and colleagues used the ADOS as a context for recording, transcribing, and coding language samples during eight activities common to Module 2 and 3 on the ADOS (Park, Yelland, Taffe, & Gray, 2012). Similarly, Kover and colleagues recorded, transcribed and coded language samples obtained during the first 15 minutes of the ADOS across Modules 1, 2, and 3 (Kover, Davidson,

Sindberg, & Weismer, 2014). However, both studies used time-intensive language sampling, which requires both specialized resources and expertise.

Some other studies have used the ADOS Module administered as a categorical indicator of verbal/nonverbal status. For example, Bal and colleagues (2016) classified children into two general categories based on the ADOS Module administered: Minimally Verbal (Module 1) and Verbal (Modules 2, 3, or 4). The Minimally Verbal group was then subcategorized into "Few-to-No Words" and "Some Words" groups based on whether they used more or less than five words (Bal et al., 2016). Thurm and colleagues used a similar approach to categorize children into two groups (Minimally Verbal or Phrase Speech) based on a combination of ADOS module and score on item A1 (Thurm, Manwaring, Swineford, & Farmer, 2015). However, neither study assessed convergent validity, specifically, nor utilized the full range of possible ADOS item-level scores focusing on spoken language. Consideration of module alone does not provide an ability to examine finer-grained differences in expressive language that may be present between individuals both within and across modules.

Interestingly, all modules of the ADOS-2 contain an item (A1) on which the examiner rates the participant's current level of spoken language. By consolidating scores on this item across modules into a single metric it would be possible to derive a single expressive language score. Such a score would be useful for behavioral phenotyping since it could be derived from a widely-used measure that is applicable across ages and functional levels. Thus, the purpose of the current study was to calculate an Expressive Language Score from the ADOS-2 that could be derived across modules and to examine the utility and convergent and divergent validity of this new score.

Methods

Participants

Participants included 470 children with ASD and their parents who were participating in a larger study examining the psychometric properties of an autism assessment tool. Participants were recruited from one of three sites: BLINDED FOR REVIEW. Eligibility criteria included a previous diagnosis of and meeting DSM-IV or 5 criteria for ASD, meeting or exceeding clinical cut-off scores on the ADOS-2 (Lord et al., 2012), and being between 2 and 14 years of age (M = 7.3 years, SD = 3.8). The sample was mostly male (82.1%) and predominantly Caucasian (81.1%) and non-Hispanic/Latino (77.7%). Full Scale IQ ranged from 30 to 141 (M = 84.6, SD = 23.5), as assessed by either the Early Learning Composite score of the Mullen Scales of Early Learning (Mullen, 1995) or the General Conceptual Ability score of the Differential Ability Scales, Second Edition (Elliot, 2007).

Measures

ADOS-2 Expressive Language Score

An expressive language score was derived from clinician ratings on the ADOS-2 (Lord et al., 2012). The ADOS-2 is a standardized, semi-structured diagnostic observational tool designed to assess communication, social interaction, repetitive behaviors, restricted interests, and other behavioral features of ASD. The ADOS-2 contains five different modules (Toddler Module and Modules 1-4). The clinician selects the most appropriate module for administration based on the individual's age and verbal ability. Each module consists of a number of different activities, ranging across modules from play-based tasks to conversation, each designed to elicit spontaneous social initiations, responses, and other behaviors. Each module includes a number of different items that are coded following the observation, with most item scores ranging from 0 (the specified abnormality is absent) to 2 or 3 (the abnormality is definitely present) in most

cases. The ADOS-2 was administered by examiners who had achieved research-reliability on scoring and administration of the measure.

In the current study, Item A1 (Overall Level of Non-Echoed Spoken Language) was used to create an ADOS expressive language score. Item A1 is included across all modules of the ADOS-2 with comparable (and mutually exclusive) coding conventions across modules. In this way, a single score was calculated by utilizing the full range of possible scores across modules, with a possible range of 1 ("no spontaneous use of words or word approximations") to 8 ("uses sentences in a largely correct fashion, must use some complex speech") (see Table 1). Scores were converted as follows (**1**= codes of "4" on MT or M1; **2** = codes of "3" on MT or M1; **3** = codes of "2" on MT or M1; **4** = codes of "1" on MT or M1 or codes of "2" on M2; **5** = codes of "0" on MT or M1 or codes of "1" on M2; **6** = codes of "0" on M2 or codes of "2" on M3 or M4; **7** = codes of "1" on M3 or M4; and **8** = codes of "0" on M3 or M4). Note that scores of "3" on M2, M3 or M4 are not included in the score calculation because they do not have a 1:1 correspondence with codes from other modules, and in some cases may be an indication that a different module would have been more appropriate.

Concurrent and Divergent Validity Measures

Additional measures were included to examine the extent to which the ADOS Expressive Language score correlated with other measures of expressive language (concurrent validity), and the extent to which it was less strongly associated with measures of theoretically distinct constructs (divergent validity). Divergent validity measures of receptive language and nonverbal cognitive ability were selected because they assess similar, yet conceptually distinct constructs. As such, associations with these measures were expected to be significantly weaker than with convergent validity measures.

Parent Report Measures

The Vineland Adaptive Behavior Scales (VABS-II) – Expressive Communication raw score (Sparrow, Cicchetti, & Balla, 2005) was included as a convergent validity measure of expressive language, and the Receptive Communication raw score was included to examine divergent validity. The Total Words Produced subscale raw score from the MacArthur Communicative Development Inventory (CDI) Words & Gestures form (Fenson et al., 1993; Fenson et al., 1994) was included as a second parent-report measure of expressive language, and the Total Words Understood subscale raw score was included to examine divergent validity. Because the CDI was not appropriate for older children with fluent language, it was only administered to a subset of the sample (n = 103, 21.9%).

Additional parent-report measures of expressive language included the Communication subscale of the Autism Impact Measure (AIM) (Kanne et al., 2014, Mazurek & Kanne, 2017) and parent-report of the child's overall general spoken language ability as assessed by responses items on a parent-report history form developed for the larger study (1 = no single words, 2 = single words only, 3 = simple phrases, and 4 = complete sentences).

Direct Assessment

Participants were administered one of two measures of cognitive functioning, depending on age and/or expressive language level: the Mullen Scales of Early Learning (MSEL; 15.3% of the sample) (Mullen, 1995) or the Differential Ability Scales, Second Edition (DAS-II) Early Years battery (39.6% of the sample) or School-Age Battery (40.6% of the sample) (Elliot, 2007). For children who were unable to obtain a basal score on the age-appropriate measure or battery, the examiner administered the next alternative measure until a basal was achieved. As a result

the MSEL was administered out-of-age-range to 18 participants, and the DAS-II Early Years battery was administered out-of-age-range to 13 participants.

For the subset of children who received the MSEL, the Expressive Language raw score was included as a concurrent measure of expressive language, and the Receptive Language and Visual Reception raw scores (assessing receptive language and nonverbal cognitive ability, respectively) were included to examine divergent validity. For children receiving the Early Years DAS-II battery, the Naming Vocabulary subtest ability score was included as a concurrent measure of expressive language. For those receiving the School-Age DAS-II battery, the Word Definitions subtest ability score was included as a measure of expressive language. The DAS-II Matrices subtest (assessing nonverbal cognitive ability) was included to examine divergent validity for both Early Years and School-Age batteries.

Clinician Rating of Expressive Language

The Verbal Communication subdomain of the Ohio Autism Clinical Global Impression Severity Scale (OACIS-S) (Butter & Mulick, 2006) was included as concurrent clinician-report measure of expressive language. The Verbal Communication domain assesses difficulties with speech, language, and conversation (as rated by clinician through a review of all available clinical information), with scores ranging from 1 to 7 and higher scores indicating greater difficulties.

Results

Regarding descriptive statistics, the ADOS-2 Expressive Language score ranged from 1 to 8 (M = 5.6, SD = 2.4). The ADOS-2 Expressive Language score was significantly correlated with age (Spearman rank-order correlation $[r_s]$ = .65, $p < .001$) and IQ (r_s = .64, $p < .001$). This

was further evaluated by examining ADOS-2 Expressive Language descriptive statistics across age groups (see Table 2).

Convergent validity of the ADOS-2 Expressive Language Score was evaluated by calculating correlation coefficients between the ADOS-2 Expressive Language score and the convergent validity measures described above. Because data from all key variables did not meet normality assumptions required for parametric statistics and because the ADOS-2 Expressive Language score was an ordinal variable, Spearman's rank-order correlation coefficient, rho, was used. Using Cohen's (1992) conventions for interpreting the effect size of correlation coefficients as small = .10, medium = .30, large = .50, as shown in Table 2, the strength of association was moderate to large across measures, providing good evidence of convergent validity and for the use of this new score. To control for potential effects of age on these associations, we also conducted partial correlations controlling for age (see Table 3), and the results indicate that all correlations remained significant with similar effect sizes. Regarding divergent validity of the ADOS-2 Expressive Language score, the results indicated significantly stronger correlations with measures of expressive language than with measures of receptive language, and significantly stronger correlations with measures of expressive language than with measures of nonverbal cognitive ability (see Table 3).

Discussion

Expressive language is an important and clinically relevant aspect of the ASD phenotype. However, many ASD research studies do not include specific measures of language functioning and do not adequately describe the verbal abilities of their participants. A valid and efficient measure of expressive language functioning in children with ASD would foster increased consistency across ASD research. The current study calculated an ADOS-2 expressive language

score and examined the convergent and divergent validity of this new language metric in a large sample of children and adolescents with ASD.

The results indicated good convergent validity of this ADOS-derived score, as evidenced by moderate to large correlations with other measures of expressive language. Specifically the ADOS-2 expressive language score was strongly associated with parent-report measures of expressive language and communication (including the Vineland, the Macarthur CDI, the AIM, and an overall estimate of spoken language). Importantly, the ADOS-2 expressive language score was also strongly correlated with direct measures, including expressive language subtests from both the Mullen Scales of Early Learning and the DAS-II. The ADOS-2 expressive language score was also significantly associated with a clinician-rated measure of verbal communication: the OACIS-S Verbal Communication score. However, it should be noted that clinicians took all clinical information into account when assigning these ratings, including a child's performance on the ADOS-2. Although the expressive language metric was not calculated or specifically considered by clinicians in assigning OACIS-S scores, item-level ADOS-2 scores were available for consideration. The pattern of results also provided support of divergent validity for this new metric, as correlations were significantly stronger with measures of expressive language than with measures of either receptive language (two highly related but separate constructs) or nonverbal cognitive ability.

While not a comprehensive assessment of language, the ADOS-derived score offers notable benefits for use in sample characterization. First, the ADOS-2 is a widely-used ASD diagnostic measure that is applicable to individuals across ages and is one of the most widely used behavioral observation tools. Secondly, it fosters efficiency by enabling brief direct language characterization in situations in which the ADOS-2 is already being administered

without need for additional time, resources, or technical expertise. However, it should be noted that this brief score should not be substituted for a comprehensive language assessment, especially during diagnostic assessments. It is also important to recognize that the ADOS-2 itself requires formal specialized training for reliable administration. Thus, while the use of this score may facilitate standard reporting of language across ASD research studies that are already using the ADOS-2, it does not necessarily represent an efficient stand-alone measure of language in situations in which the ADOS-2 is not being administered for other purposes.

The current results also suggest that this new score may be most useful for characterizing expressive language for younger children and/or those with less well-developed language abilities. The ADOS-2 item scores offer more fine-grained differentiation of abilities on the Toddler Modules and Modules 1 and 2 than on Modules 3 and 4. Thus the derived metric will likely be most sensitive to differences for children before they acquire complex speech.

The current results suggest that the ADOS-2 expressive language score may provide a simple and efficient measure of verbal abilities. However, future research is needed to further validate this score. Although the sample represented a broad range of age and functional level, it is somewhat limited in terms of racial and ethnic diversity and the study did not include individuals who were administered Module 4 of the ADOS-2. The current study did not include comprehensive measures of language functioning or natural language samples. Thus, future studies should examine whether the ADOS-2 expressive language score correlates with additional direct measures of language, such as the Clinical Evaluation of Language Fundamentals (CELF) (Wiig, Semel, & Secord, 2003; Wiig, Secord, & Semel, 2004), the Preschool Language Scales (PLS) (Zimmerman, Steiner, & Pond, 2011), and natural language sampling across contexts and interaction partners. It would also be informative to determine

whether this score can be used to accurately characterize expressive language in typically developing children and those with other developmental disorders. If so, this score may be potentially useful for briefly characterizing (or matching) language abilities of children without ASD who are enrolled in ASD research studies within control or comparison groups and who are also administered the ADOS-2 assessment tool. Finally, future research should evaluate the extent to which this new ADOS-2 expressive language score may be useful in tracking change over time. Overall, the results of this study suggest that the new ADOS-2 expressive language score demonstrates good convergent and divergent validity and may offer a simple and efficient option for reporting verbal abilities across ASD research studies.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, fifth edition (DSM-5)*. Washington, DC: American Psychiatric Association.

Bal, V. H., Katz, T., Bishop, S. L., & Krasileva, K. (2016). Understanding definitions of minimally verbal across instruments: Evidence for subgroups within minimally verbal children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 57*(12), 1424-1433.

Bishop, D. V. M. (2003). *The Children's Communication Checklist-Second Edition CCC-2 Manual*. London: Psychological Corporation.

Bornstein, M. H., & Haynes, O. M. (1998). Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Development, 69*(3), 654-671.

Butter, E., & Mulick, J. (2006). The *Ohio Autism Clinical Impressions Scale (OACIS).* Columbus, OH: Children's Research Institute.

Carrow-Woolfolk, E. (2017). *Comprehensive assessment of spoken language, second edition (CASL-2) [Manual]*. Torrance, CA: Western Psychological Services.

Charman, T. (2004). Matching preschool children with autism spectrum disorders and comparison children for language ability: Methodological challenges. *Journal of Autism and Developmental Disorders, 34*(1), 59-64.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155.

Cole, K. N., Mills, P. E., & Dale, P. S. (1989). Examination of test-retest and split-half reliability for measures derived from language samples of young handicapped children. *Language, Speech, and Hearing Services in Schools, 20*, 259-268.

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology, 12*(3), 349-358.

Eigsti, I., Bennetto, L., & Dadlani, M. B. (2007). Beyond pragmatics: Morphosyntactic development in autism. *Journal of Autism and Developmental Disorders, 37*(6), 1007-1023.

Elliot, C. (2007). *Differential ability scales-II.* San Antonio, TX: The Psychological Corporation Harcourt Brace Janovich.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., . . . Reilly, J. S. (1993). *MacArthur Communicative Development Inventories: User's guide and technical manual.*

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development,* i-185.

Hale, C. M., & Tager-Flusberg, H. (2005). Brief report: The relationship between discourse deficits and autism symptomatology. *Journal of Autism and Developmental Disorders, 35*(4), 519-524.

Heilmann, J., Nockerts, A., & Miller, J. F. (2010). Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools, 41*(4), 393-404.

Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: Implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 33*(1), 3-13.

Howlin, P., Savage, S., Moss, P., Tempier, A., & Rutter, M. (2014). Cognitive and language skills in adults with autism: A 40-year follow-up. *Journal of Child Psychology and Psychiatry, 55*(1), 49-58.

Kanne, S. M., Mazurek, M. O., Sikora, D., Bellando, J., Branum-Martin, L., Handen, B., . . . Warren, Z. (2014). The autism impact measure (AIM): Initial development of a new tool for treatment outcome measurement. *Journal of Autism and Developmental Disorders, 44*(1), 168-179.

Koegel, L. K., Koegel, R. L., & Smith, A. (1997). Variables related to differences in standardized test outcomes for children with autism. *Journal of Autism and Developmental Disorders, 27*(3), 233-243.

Kover, S. T., Davidson, M. M., Sindberg, H. A., & Weismer, S. E. (2014). Use of the ADOS for assessing spontaneous expressive language in young children with ASD: A comparison of sampling contexts. *Journal of Speech, Language, and Hearing Research, 57*(6), 2221-2233.

Kwok, E. Y., Brown, H. M., Smyth, R. E., & Cardy, J. O. (2015). Meta-analysis of receptive and expressive language skills in autism spectrum disorder. *Research in Autism Spectrum Disorders, 9*, 202-222.

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,*

Lund, N. J., & Duchan, J. F. (1993). *Assessing children's language in naturalistic contexts.* Prentice Hall.

Mazurek, M.O. & Kanne, S. (2017). *Autism Impact Measure (AIM) Scoring Instructions.* Prepublication material for the AIM copyright © 2017 by Western Psychological Services.

Mayo, J., Chlebowski, C., Fein, D. A., & Eigsti, I. (2013). Age of first words predicts cognitive ability and adaptive skills in children with ASD. *Journal of Autism and Developmental Disorders, 43*(2), 253-264.

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

Park, C. J., Yelland, G. W., Taffe, J. R., & Gray, K. M. (2012). Morphological and syntactic skills in language samples of pre school aged children with autism: Atypical development? *International Journal of Speech-Language Pathology, 14*(2), 95-108.

Pickles, A., Anderson, D. K., & Lord, C. (2014). Heterogeneity and plasticity in the development of language: A 17-year follow-up of children referred early for possible autism. *Journal of Child Psychology and Psychiatry, 55*(12), 1354-1362.

Rapin, I., Dunn, M. A., Allen, D. A., Stevens, M. C., & Fein, D. (2009). Subtypes of language disorders in school-age children with autism. *Developmental Neuropsychology, 34*(1), 66-84.

Skwerer, D., Jordan, S. E., Brukilacchio, B. H., & Tager-Flusberg, H. (2016). Comparing methods for assessing receptive language skills in minimally verbal children and adolescents with autism spectrum disorders. *Autism, 20*(5), 591-604.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales (2nd ed.)*. Circle Pines, MN: American Guidance Service.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245-251.

Stiles, J. (1994). On the nature of informant judgements in inventory measures: And so what is it you want to know ? *Monographs of the Society for Research in Child Development, 59,* 180–185.

Swensen, L. D., Kelley, E., Fein, D., & Naigles, L. R. (2007). Processes of language acquisition in children with autism: Evidence from preferential looking. *Child Development, 78*(2), 542-557.

Tager-Flusberg, H. (2000). The challenge of studying language development in children with autism. *Methods for Studying Language Production,* 313-332.

Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clinical Neuroscience Research, 6*(3), 219-224.

Tager-Flusberg, H., Calkins, S., Nolin, T., Baumberger, T., Anderson, M., & Chadwick-Dias, A. (1990). A longitudinal study of language acquisition in autistic and down syndrome children. *Journal of Autism and Developmental Disorders, 20*(1), 1-21.

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., ... & Yoder, P. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research, 52*(3), 643-652.

Thomas, J. A. (1989). A standardized method for collecting and analyzing language samples of preschool and primary children in the public schools. *Language, Speech, and Hearing Services in Schools, 20*(1), 85-92.

Thurm, A., Manwaring, S. S., 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.

Ukrainetz, T. A., & Blomquist, C. (2002). The criterion validity of four vocabulary tests compared with a language sample. *Child Language Teaching and Therapy, 18*(1), 59-78.

Weismer, S. E., Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism spectrum compared to toddlers with developmental delay. *Journal of Autism and Developmental Disorders, 40*(10), 1259-1273.

Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders, 34*(5), 473-493.

Wiig, E. H., Secord, W., & Semel, E. M. (2004). *CELF preschool 2: Clinical evaluation of language fundamentals preschool.* Pearson/PsychCorp.

Wiig, E. H., Semel, E. M., & Secord, W. (2003). *CELF 5: Clinical evaluation of language fundamentals.* Pearson/PsychCorp.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *PLS-5: Preschool language scales.* Pearson/PsychCorp.

Received 2/13/2018, accepted 1/11/2019

Table 1: ADOS-2 Expressive Language Score Calculation

New Expressive Language Code	Original A1 Item Code by Module					
	Toddler	**Module 1**	**Module 2**	**Module 3**	**Module 4**	**Description**
1	4	4				No spontaneous words or word approximations
2	3	3				At least one word or word approximation, but fewer than 5 words during evaluation
3	2	2				Recognizable single words or word approximations only; must use at least 5 words during evaluation
4	1	1	2			Occasional phrases only, mostly single words
5	0	0	1			Regular use of utterances with two or more words
6			0	2	2	Non-echoed speech is mostly utterances of at least three words, but without complex language
7				1	1	Some relatively complex speech but with recurrent grammatical errors not associated with use of dialect

8				0	0	Uses sentences in a largely correct fashion (must use some complex speech)

Note: Item scores of “3” on Modules 2, 3 and 4 are not included in the score calculation because they do not have a 1:1 correspondence with codes from other modules, and in some cases may be an indication that a different module would have been more appropriate.

Table 2: Descriptive Statistics of ADOS-2 Expressive Language (EL) Scores across Age Groups

	Total Sample	**Ages 2-4**	**Ages 5-7**	**Ages 8-10**	**Ages 11-14**
	(*n* = 464)	(*n* = 179)	(*n* = 98)	(*n* = 82)	(*n* = 105)
ADOS-2 EL Score	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
	5.6 (2.4)	4.0 (1.8)	5.7 (2.4)	6.9 (2.0)	7.3 (1.6)
Frequency of ADOS-2 EL Scores within Age Group	n (%)	n (%)	n (%)	n (%)	n (%)
1	28 (6.0)	15 (8.4)	8 (8.2)	3 (3.7)	2 (1.9)
2	47 (10.1)	31 (17.3)	9 (9.2)	5 (6.1)	2 (1.9)
3	33 (7.1)	25 (14.0)	6 (6.1)	1 (1.2)	1 (1.0)
4	50 (10.8)	37 (20.7)	7 (7.1)	2 (2.4)	4 (3.8)
5	36 (7.8)	25 (14.0)	8 (8.2)	1 (1.2)	2 (1.9)
6	55 (11.9)	37 (20.7)	10 (10.2)	5 (6.1)	3 (2.9)
7	38 (8.2)	6 (3.4)	14 (14.3)	9 (11.0	9 (8.6)
8	177 (38.1)	3 (1.7)	36 (36.7)	56 (68.3)	82 (78.1)

Table 3: Correlations between ADOS-2 Expressive Language (EL) Score and Convergent and Divergent Measures[a]

		Sample size	Age *M (SD)*, range	ADOS-2 EL *M (SD)*, range	*Spearman Rank-Order Correlation* (r_s)	*Partial* r_s (Controlling for Age)	*Z*
Parent-Report							
Convergent Measures	***Divergent Measures***						
Vineland-II Expressive Communication Raw Score		*n* = 401	7.5 (3.9), 2-14	5.7 (2.4), 1-8	0.89***	0.80***	
	Vineland-II Receptive Communication Raw Score				*0.61****		*14.0****

CDI Total Words Produced Subscale		*n* = 103	3.8 (1.7), 2-11	2.8 (1.4), 1-8	0.84***	0.84***	
	CDI Total Words Understood Subscale				*0.67****		*4.4****
AIM Communication Subscale[b]		*n* = 450	7.3 (3.8), 2-14	5.6 (2.4), 1-8	-0.68***	-0.58***	
General Spoken Language Level		*n* = 459	7.3 (3.8), 2-14	5.6 (2.4), 1-8	0.75***	0.65***	
Direct Assessment							
Convergent Measures	***Divergent Measures***						
MSEL Expressive Language Raw Score		*n* = 72	4.6 (3.1), 2-14	2.4 (1.2), 1-5	0.74***	0.74***	
	MSEL Expressive Language Raw Score				*0.47****		*4.1****

	MSEL Visual Reception Raw Score				*0.39***		*6.1****
DAS-II Early Years Naming Vocabulary Ability Score		*n* = 173	4.7 (1.6), 2-14	5.0 (1.8), 1-8	0.80***	0.79***	
	DAS-II Early Years Matrices Ability Score				*0.54****		*4.9****
DAS-II School-Age Word Definitions Ability Score		*n* = 185	11.1 (2.3), 5-14	7.7 (0.9), 2-8	0.51***	0.57***	
	DAS-II School-Age Matrices Ability Score				*0.20**		*5.1****
Clinician Ratings							

Convergent Measure					
OACIS-S Verbal Communication Subdomain[b]	*n* = 457	7.3 (3.8), 2-14	5.6 (2.4), 1-8	-0.47***	-.59***

* $p < .01$; ** $p < .001$; *** $p < .0001$

[a]Due to differences in sample size across measures, sample size is reported for each analysis and descriptive statistics for age and ADOS-2 EL score are reported separately by concurrent measure. Differences in correlations between ADOS EL scores and convergent versus divergent measures were examined via Z-test for comparison of dependent correlations (Steiger, 1980).

[b]Note that higher scores on the AIM and OACIS-S Communication subscales indicate greater difficulties with communication.