Title
Using weighted communication scoring procedures in naturalistic play samples: Preliminary validation in preschool-aged boys with autism or fragile X syndrome

Journal
Autism Research, 15(9)

ISSN
1939-3792

Authors
Thurman, Angela John
Alvarez, Cesar Hoyos
Nguyen, Vivian

Publication Date
2022-09-01

DOI
10.1002/aur.2724

Peer reviewed

HHS Public Access
Author manuscript
Autism Res. Author manuscript; available in PMC 2023 September 01.

Published in final edited form as:
Autism Res. 2022 September ; 15(9): 1755–1767. doi:10.1002/aur.2724.

Using Weighted Communication Scoring Procedures in Naturalistic Play Samples: Preliminary Validation in Preschool-Aged Boys with Autism or Fragile X Syndrome

Angela John Thurman[a,b], **Cesar Hoyos Alvarez**[b,c], **Vivian Nguyen**[a,b]

[a]Department of Psychiatry and Behavioral Sciences, University of California Davis Health

[b]MIND Institute, University of California Davis Health

[c]Department of Spanish and Portuguese, University of California Davis

Abstract

In the present study, we provide a preliminary evaluation of the validity and reliability of using weighting procedures to measure communication in play samples for preschool-aged boys with autism or fragile X syndrome (FXS). Because weighting procedure communication scores (WPCSs) reflect growth in both communicative frequency and complexity, establishing the psychometrics of the component scores, in addition to the overall metric, affords investigators the opportunity to describe growth within and across skills. Results, for both groups, provide support regarding the psychometric appropriateness (i.e., convergent validity, divergent validity, and internal consistency) for all WPCSs. That said, a trend was observed for reliability scores to be slightly lower or more variable in boys with autism than in boys with FXS. Finally, although significant associations were observed in the associations between WPCSs across play contexts, contexts effects were observed for all three WPCSs. Together, results from this study provide promising preliminary data indicating the utility of using WPCSs in children with neurodevelopmental disabilities.

Lay Summary:

Language supports long-term positive outcomes; it is important to identify accurate and flexible ways of measuring language in children over time. We considered the effectiveness of using a procedure that considers changes in the number of communication acts and the types of acts produced during a play session by preschool-aged boys with autism or fragile X syndrome. These procedures were found to be valid and reliable.

Keywords

communication; outcome measure; weighting communication scores; naturalistic; fragile X syndrome; autism

Corresponding Author: Angela John Thurman ajthurman@ucdavis.edu.

Introduction

Language supports a wide range of positive outcomes, including development in other neurocognitive systems, socialization skills, and academic achievements (Clegg, Hollis, Mawhood, & Rutter, 2005; Hartley et al., 2011; Homer & Tamis-LeMonda, 2005). Compared to the rapid and predictable rates of language acquisition observed in children with typical development (TD), children with neurodevelopmental disabilities (NDDs), such as autism or fragile X syndrome (FXS), often demonstrate more delayed and variable language trajectories (Abbeduto, McDuffie, Thurman, & Kover, 2016). These differences make measuring early communication more difficult; in fact, a lack of outcome measures validated for use in NDDs is a significant barrier for descriptive and treatment research (Berry-Kravis et al., 2018; Kelleher & Wheeler, 2020). The present study provides a preliminary evaluation of the validity and reliability of weighting procedures to measure expressive communication during naturalistic play samples in preschool aged boys with autism or FXS. This approach allows the consideration of communication within and across different developmental stages (e.g., prelinguistic, word combinations).

Challenges Associated with Measuring Expressive Communication Skills

Investigations focused on describing and supporting language development require valid and reliable assessment tools (Berry-Kravis et al., 2018; Kelleher & Wheeler, 2020). However, many of the tools available have serious weaknesses (Berry-Kravis et al., 2018). The need for such tools is particularly acute during early childhood because early interventions can create a cascade of long-term positive events (D'Souza & Karmiloff-Smith, 2017).

Multiple challenges are encountered when assessing early language development. One challenge involves the limited number of items on standardized assessments dedicated to the transition between the production of first words and word combinations. Consider both the Preschool Language Scales-5 (PLS-5; Zimmerman, Steiner, & Pond, 2011), a comprehensive language assessment of preverbal skills through early literacy, and the Mullen Scales of Early Learning (Mullen, 1995), an assessment of cognitive development. Only 12/64 PLS-5 expressive communication items index skills between "use of one word" and "uses different word combinations," and only 6/28 Mullen expressive language items index skills between "says first words" and "uses two-word phrases." The limited number of items dedicated to this developmental period reflects the fact that children with TD transition between first words and word combinations rapidly between the first and second years of life (Bloom & Capatides, 1987; Braine & Bowerman, 1976; Fenson et al., 2007). For children with NDDs, however, many will only begin to enter this stage during their preschool years (Abbeduto et al., 2016). There is also considerable variability across children with NDDs in the ages at which they transition both in and out of this stage. Because of these differences, tools that can be applied across a wider chronological age period and are more sensitive to developmental improvements are needed.

Another challenge to assessing early language development is that indicators of progress vary across developmental stages. For example, during the prelinguistic stage, children transition across five vocal stages, beginning with the production of reflexive sounds, then partially resonant vowel productions, fully resonant vowel productions, consonant-vowel combinations, and finally, the production of meaningful speech (Kuhl & Meltzoff, 1998). Thus, communicative growth can be measured by monitoring children's vocal development across these stages. Similarly, gestural skills emerge before the end of the first year of life and predict their word forms, even in children who use words as their primary mode of communication (Capone & McGregor, 2004; Iverson & Goldin-Meadow, 2005). Thus, changes in the number and types of gestures used can also be used to indicate communicative growth. Finally, prelinguistic skills, such as vocalizations, often steadily increase and then gradually decrease as children shift toward linguistic behaviors (Greenwood et al., 2013). Thus, measures that consider growth within and across language stages are needed when working with children with NDDs who demonstrate considerable variability in their communication skills or when tracking development across language stages.

Weighting Procedures

One technique used to navigate the complexity of tracking growth within and across skill levels is to create weighted composite scores. First, skills are classified using a developmental hierarchy. Next, the frequency of each skill is weighted, with weights increasing along the developmental hierarchy. For example, in the Early Communication Indicator, Greenwood et al. (2006) considered the occurrences of vocalizations, gestures, single-word utterances, and multiple-word utterances in children. Each occurrence of a prelinguistic form (i.e., vocalizations and gestures) received 1 point; each single-word utterance received 2 points; and each multiple-word utterance received 3 points. These weighted scores were then summed to generate a weighted communication score that increased steadily until 3.5 years of age (Greenwood et al., 2006). Similarly, in the Communication Complexity Scale, a communication measure for use in individuals with NDDs who are primarily prelinguistic communicators, skills are weighted along a 12-point developmental continuum marking development from pre-intentional to intentional symbolic communication (Brady et al., 2018). Mazurek and colleagues also used this weighting approach when proposing the new ADOS-2 expressive language score (Mazurek, Baker-Ericzén, & Kanne, 2019). Across all these scales, the use of weighting procedures allows for transitions across important milestones/developmental stages to be monitored, which closely aligns with practices for developing intervention goals (Tager-Flusberg et al., 2009).

Weighting procedures also offer flexibility for use in various methodological designs. For example, weighting procedures have been used to assess the communication skills of children with autism across a variety of contexts, including contexts designed to press for symptoms associated with autism (Boyd et al., 2018; Yoder, Stone, & Edmunds, 2020; Yoder, Stone, Walden, & Malesa, 2009), semi-structured naturalistic communication samples with examiners (Yoder et al., 2020), and free play contexts with parents (Carter et al., 2011). Weighting procedures have also been used to describe the development of prelinguistic skills (Boyd et al., 2018), language skills (Rosenberg & Abbeduto, 1987; Thurman & Hoyos Alvarez, 2020), and both prelinguistic and language skills simultaneously (Greenwood et al., 2006; Yoder et al., 2009).

Unfortunately, the psychometric properties of many weighting procedures have not been evaluated for use when characterizing behavior in naturalistic communication samples, which are commonly used to assess language development (Abbeduto et al., 2020). Moreover, these procedures are argued to be a promising alternative to standardized tests because the contexts are more closely aligned with the contexts in which children naturally communicate (Westerveld, Gillon, & Miller, 2004). Psychometric data on using weighting procedure communication scores (WPCSs) in naturalistic play samples are needed to establish their appropriateness for use in treatment and other research studies.

Comparisons of Language Skills in Boys with Autism or FXS

Language delays are commonly observed in children with NDDs, with varying patterns of strength and challenge, and considerable heterogeneity observed both within and across conditions (Abbeduto et al., 2016). Consider the autism and FXS phenotypes. Autism is a multifactorial, behaviorally defined condition (American Psychiatric Association, 2013; Ronald & Hoekstra, 2011); FXS results from the silencing of a single gene on the X chromosome (the *FMR1* gene; Bassell & Warren, 2008). Both conditions are observed more commonly in males than in females, but in FXS, the behavioral features associated with the phenotype are also more pronounced in males than in females (Rinehart, Cornish, & Tonge, 2010).

Both of these phenotypes are associated with increased risk for language delays, albeit to varying degrees depending on what factors are considered within comparisons. As a group, boys with FXS typically demonstrate more significant language delays than do same-aged boys with autism (Kjelgaard & Tager-Flusberg, 2001; McDuffie, Thurman, Channell, & Abbeduto, 2017), in part due to the differential rates of intellectual disability between FXS (Hessl et al., 2009) and autism (Maenner et al., 2020). That said, there is evidence that boys with FXS may demonstrate a strength in language skills relative to boys with autism, particularly after controlling for differences in nonverbal IQ and autism symptom severity (Sterling, 2018; Thurman & Hoyos Alvarez, 2020; Thurman, McDuffie, Hagerman, Josol, & Abbeduto, 2017). Because of the impact language development has on adaptive outcomes, monitoring language development in these conditions is of critical interest for both research and clinical purposes (Tager-Flusberg et al., 2009).

Present Study

In this study, we conducted a preliminary investigation of the psychometric properties of WPCSs in preschool-aged boys with autism or FXS, two NDDs associated with varying degrees of language delay relative to developmental expectations. The following research questions were addressed:

1. Can construct validity (i.e., convergent validity, discriminant validity, and replication of prior findings of between-group differences) be established for WPCSs?
2. Is adequate internal consistency (i.e., split-half reliability and reliability across play activities) observed for WPCSs?
3. Do WPCSs vary across different play activities?

Methods

Participants

Participants included 28 boys with autism (*M*(CA) = 4.43, *SD* = .77) and 21 boys with FXS (*M*(CA) = 4.20, *SD* = .83) between 3.00 and 5.50 years of age, drawn from a longitudinal study examining early language abilities in children with autism or FXS. Descriptive statistics for the sample are presented in Table 1. Families of participants with autism provided documentation of an existing community diagnosis of autism spectrum disorder, confirmed through administration of the Autism Diagnostic Observation Schedule-2 (ADOS-2; Lord et al., 2012). Families of participants with FXS provided documentation of a diagnosis of the *FMR1* full mutation (i.e., >200 CGG repeats, with or without mosaicism). In the present study, a caregiver participated in a play sample with the child. The ethnic/racial composition of the participants with autism or FXS, and of their participating caregivers, are presented in Table 2.

Participants were recruited through the MIND Institute's IDDRC Clinical Translational Core Registry, parent listservs, the National Fragile X Foundation, and through clinics and preschools specialized in working with children with NDDs. Due to the higher prevalence rate, participants with autism were more likely to reside locally than were those with FXS. All participants enrolled in the present study met the following criteria (based on parent report): (a) boys ages 36–66 months at enrollment, (b) English is the primary language of exposure; (e) no sensory or physical impairments that would limit participation in project activities; and (f) no medical conditions (e.g., severe and frequent seizures) that prevented them from meeting the demands of the testing protocol. In addition, participants with autism with above-average intellectual ability (parent report or prior documentation of IQ greater than 110 or greater) were not included in the present project since nearly all boys with FXS demonstrate cognitive skills in the low average to intellectual disability range. Approval from the Institutional Review Board, as well as parental informed consent, was obtained. All assessments took place in the research laboratory and were conducted by PhD-level study personnel or trainees under their supervision.

Measures

Weighting procedure communication scores in naturalistic play samples.

Play Sample.: The Abbreviated-Communication Play Protocol (CPP-A), a 20-minute semi-naturalistic caregiver-child play session, was completed (Adamson, Bakeman, Deckner, & Romski, 2009). The CPP-A uses the metaphor of a "play" with different "scenes" to guide the caregiver through multiple 5-minute play activities. These different "scenes" (hereafter referred to as activities) were designed to systematically sample a range of communicative functions (i.e., requesting, commenting, and social interaction), providing researchers an opportunity to collect a single sample that considers communication across a range of communicative functions or to look at a specific communicative function (Adamson & Bakeman, 2016). Once the caregiver and child were comfortable and ready to begin, the research explained the play sampling procedures. At the start of each 5-minute activity, the researcher set up the activity toys, explained the focus of the activity, and provided the parent with a cue card that could be used as a reference regarding the activity and directional suggestions to encourage child communication. Four 5-minute activities were completed:

- Free Play: The parent was encouraged to play as they normally would.
- Social Interaction: The parent was encouraged to engage the child in a turn-taking routine
- Requesting: Toys were placed on a shelf out of the child's reach and the parent was encouraged to draw the child's attention to the toy and help the child access the toy.
- Commenting: The parent was provided with toys in an opaque container and encouraged to share objects from the container one at a time.

Once the parent was comfortable proceeding with the activity, the researcher collected the toys from the previous scene and said goodbye to the child and parent before exiting (see Table 3 for toys and directional suggestions). The CPP-A has been a useful tool for measuring communication skills in children with TD or NDDs; additional detailed information on the CPP-A is provided by Adamson and Bakeman (2016).

Weighting Procedure Communication Scores.: Samples were video-recorded and coded using Behavioral Observation Research Interactive Software (Friard & Gamba, 2016). During initial training, coders achieved 85% agreement on three consecutive samples. During each 5-minute activity, every instance of four key skill types were coded.

- Gestures (weight = 1): a directed, communicative, manual movement made by the child to express an idea or meaning. Examples include deictic (e.g., give, point, show), conventional (e.g., thumbs up, wave), and descriptive (e.g., using hands to demonstrate crash) gestures. Procedures outlined by Prizant & Wetherby (2001) regarding how to determine whether an act was directed toward an adult (e.g., touching or moving toward the adult, paring with eye contact or vocalization) were used.
- Consonant-vowel vocalizations (weight = 1): a non-word or unintelligible utterance that included both a consonant and vowel (e.g., da, gaga) and was directed to the caregiver (Prizant & Wetherby, 2001) or produced during a joint engagement state (i.e., the child and caregiver were jointly attending to the object/event).
- Single-word utterance (weight = 2): Single-word intelligible utterance, voiced or manually signed.
- Multiple-word utterances (weight = 3): Multiple-word intelligible utterance, voiced or manually signed. Multiple-word utterances formed by combining both modalities were included (e.g., manually signed "more" + voiced car).

Utterances were segmented into Communication units (C-units), which has been shown to demonstrate a more accurate measure of linguistic/grammatical ability (detailed information provided by Abbeduto et al., 2020 and available from authors upon request). A C-Unit consists of an independent clause and any modifiers (e.g., dependent clauses and phrases).

An independent clause is any unit of spoken language that can stand on its own without a loss of meaning (e.g. The boy cried.). A dependent clause or phrase cannot stand on its own and requires an independent clause to give it full meaning (e.g. The boy cried [on his mother's shoulder]). Additionally, two main clauses that appear in the same sentence and are linked by coordinating conjunctions, such as and, are separated into two C-Units (e.g., He climbed up the tree. And then he jumped down.). Sentence fragments and elliptical responses also constitute C-units (e.g., More cookie.).

Three different WPCSs were generated: (1) Weighted Communication Score: the sum of the weighted-frequency scores; (2) Unweighted Communication Score: the sum of the unweighted frequency scores (i.e., each skill = 1 point); (3) Mean Weighted Communication Score: the Weighted Communication Score divided by the Unweighted Communication Score. Reliability was randomly assessed for 12% of the study sample (6% autism, 6% FXS). Intraclass Correlation Coefficient (ICC) estimates and their 95% confidence intervals were computed. A mean-rating (k = 2), absolute-agreement, 2-way mixed-effects model was computed. Results indicate excellent reliability for the metrics considered in analyses: Weighted Communication Score (ICC = 0.99; 95% confidence interval, 0.96–1.00), Unweighted Communication Score (ICC = 0.99; 95% confidence interval, 0.88–1.00), and Mean Weighted Communication Score (ICC = 0.99; 95% confidence interval, 0.99–1.00).

Regression control and participant descriptive measures.

Autism Diagnostic Observation Scale-2 (ADOS-2; Lord et al., 2012).: The ADOS-2 was administered by PhD-level researchers trained to research reliability standards. The Comparison Score from the ADOS-2, an indicator of overall autism severity level, was used in regression analyses to compare WPCSs between the diagnostic groups. 34 Module 1s (autism = 20, FXS = 14) and 13 Module 2s (autism = 8, FXS = 5) were administered. Two ADOSs were missing in the FXS sample due to scheduling difficulties. In addition, the Expressive Language Score described by Mazurek et al. (2019), which provides information regarding participant language levels, was used descriptively to characterize our participant samples and is presented in Table 1.

Differential Ability Scales-II (DAS-II; Elliott, 2007).: The Special Nonverbal Composite (SNC) from the DAS-II, which reflects nonverbal cognition, in regression analyses to compare WPCSs between the diagnostic groups. Participants under 3.5 years of age receive 2 core subtests measuring nonverbal ability (picture similarities and pattern construction). Participants 3.5 years of age and older receive 4 core subtests measuring nonverbal ability (picture similarities, matrices, pattern construction, and copying). The general population SNC mean is 100 (*SD* = 15). In addition, DAS-II Verbal SS, which reflects verbal cognition, was used descriptively to characterize our participant samples and is presented in Table 1.

Convergent validity measures.—Convergent validity is established by considering the extent to which tests of similar constructs are correlated with one another. Multiple assessments of language performance were considered in relation to WPCSs to assess convergent validity.

Differential Ability Scales-II (DAS-II; Elliott, 2007) – Verbal Subtests.: The DAS-II Verbal subtests, Verbal Comprehension and Expressive Vocabulary, were considered in study analyses. Ability scores, similar to growth scores, from both subtests, were used as validation measures.

Peabody Picture Vocabulary Test – 4th Edition (PPVT-4; Dunn & Dunn, 2007).: Growth scores from the PPVT-4, a measure of receptive vocabulary, were used in analyses. Administration of two parallel forms, version A and version B, was alternated across participants in each group.

Expressive Vocabulary Test – 2nd Edition (EVT-2; Williams, 2007).: Growth scores from the EVT-2, a measure of expressive vocabulary, were used in analyses. Administration of two parallel forms, version A and version B, was alternated across participants in each group.

Discriminant validity measures.—Discriminant validity is established by comparing measures of constructs theoretically dissimilar to one another to ensure that they demonstrate little to no association with one another. Caregiver report measures of attentional difficulties and anxiety symptoms were used to assess discriminant validity.

ADHD Rating Scale-IV Preschool Version (ADHD-RS-PV; McGoey, DuPaul, Haley, & Shelton, 2007).: The Total score from the ADHD-RS-PV, a caregiver questionnaire that assesses ADHD symptomatology, was used to assess discriminant validity. Caregiver respondents rate each item on a Likert scale of 0 (not at all) to 3 (very often).

Revised Preschool Anxiety Scale (PAS-R; Edwards, Rapee, Kennedy, & Spence, 2010).: The Total score from the PAS-R, a caregiver questionnaire that assesses anxiety symptomatology, was used to assess discriminant validity. Caregiver respondents rate each item on a Likert scale of 0 (not true at all) to 4 (very often true).

Analysis plan.—Although review of descriptive analyses (i.e., kurtosis, skew, and Kolmogorov-Smirnov tests) indicated little concern of non-normality, visual inspection identified some clustering at the distribution ends. A conservative approach was taken, with analyses considering both using zero-order bivariate correlations and Spearman rank-order correlations. Consistent findings were observed across all comparisons, with only minor variations in correlation strength observed across the two methods. We present the Spearman rank-order correlations in our results for both convergent validity and divergent validity. Bland Altman plots, and associated analyses, were also conducted to consider agreement between WPCSs and validations measures. As additional support for construct validity, we considered whether prior findings of between-group differences in language skills (one-tailed tests were used) could be replicated, using a series of multiple regression analyses to compare WPCSs between the two groups, after controlling for age, nonverbal IQ, and autism symptom severity. Models were inspected to ensure the appropriateness of regression analyses. To assess split-half reliability, we divided each communication sample into two equal halves, by alternating the assignment of each 75 second interval to version A and version B. In addition, internal consistency was assessed by considering the associations between WPCSs across different play activities utilized within the play sample. Cronbach's alpha (α) was computed to assess reliability in both analyses. Finally, mixed ANOVAs were conducted to determine if WPCSs varied as a function of the different play activities. Statistical assumptions were checked. The assumption of sphericity was violated in the Unweighted Communication Score model; a Greenhouse-Geiser correction was applied. All other assumptions were met. In all analyses, corrections for multiple comparisons were made using Benjamini and Hochberg's (1995) False Discovery Rate (FDR) procedures to maintain a familywise rate of $p \leq 0.05$.

Results

Construct Validity

For both participant groups, significant associations were observed between all WPCSs and validation measures after applying the FDR correction (see Table 4). Bland Altman plots were used to compare the two measurement techniques (i.e., see Figure 1 for Weighted Communication Score Bland Altman plots for example). No significant differences were observed between the z-scores differences between WPCSs and validation measure comparisons (*p:* .67 – 1.0). Considerations of proportional bias were not significant across all comparisons (p = .70 – 1.0). Regarding discriminant validity, for both diagnostic groups, the associations between WPCSs and the discriminant validity measures close to zero (autism: *rs:* −.02 to −.14, *ps*: .66 to .99; FXS: *r*s: −.14 to .08, *ps*: .45 to .79). Prior findings within the literature indicating that higher communication scores are observed for boys with FXS, than for boys with autism, after controlling for the effects of age, nonverbal IQ, and overall autism symptom severity, were also replicated using WPCSs (Table 5). Boys with FXS were observed to achieve a significantly higher Weighted Communication Score (154.73 points higher) and Unweighted Communication Score (68.17 points higher) than did boys with autism, after controlling for age, nonverbal cognitive ability, and autism symptom severity. These differences remained significant after the application of the FDR. Diagnostic group was not a significant independent predictor of Mean Weighted Communication Score (0.14 points higher in FXS).

Internal Consistency

High split-half reliabilities were observed in both diagnostic groups for the Weighted Communication Score (autism: α = .97, r = .95; FXS: α = .98, r = .97), Unweighted Communication Score (autism: α = .96, r = .92; FXS: α = .97, r = .95), and the Mean Weighted Communication Score (autism: α = .96, r = .92; FXS: α = .98, r = .97). Across activities, high internal consistency scores were observed for boys with FXS for all WPCSs (Weighted Communication Score α = .94; Unweighted Communication Score α = .90; Mean Weighted Communication Score α = .98). For boys with autism, scores were slightly more variable, with strong reliability scores observed for the Weighted Communication Score (α = .85) and the Mean Weighted Communication Score (α = .91). Slightly lower, albeit still acceptable, internal consistency scores were observed for the Unweighted Communication Score (α = .80) in the autism group. For both groups, significant associations across all play activities remained after applying the FDR correction, with the exception of the association between the Unweighted Communication Score in the requesting and social interaction activities for boys with autism (see Table 6).

Play Activity Comparisons

A significant effect of play activity for all WPCSs was observed (Weighted Communication Score: $F(3,141) = 5.29$, $p = .002$, $\eta_p^2 = .10$; Unweighted Communication Score: $F(2.40,112.72) = 5.38$, $p = .004$, $\eta_p^2 = .10$; and Mean Weighted Communication Score $F(3,135) = 3.33$, $p = .02$, $\eta_p^2 = .07$). Across play activities (see Figure 2), both the Weighted Communication Score *(M(difference):* 13.29 – 25.96) and the Unweighted Communication Score (*M(difference):* 5.39 – 12.61) were significantly lower in the Commenting Activity when compared to all other activities; the Unweighted Communication Score was also significantly higher in the Requesting Activity (*M(difference):* 7.21) than the Free Play Activity. The Mean Weighted Communication score in the Social Interaction activity was significantly lower than all other scores (*M(difference):* .09 – .11).

Discussion

Monitoring language development is critical for research and clinical purposes (Tager-Flusberg et al., 2009). Children with NDDs, such as autism or FXS, often demonstrate delays or more variable language trajectories than do children with TD; this makes measuring communication development more challenging. WPCSs can help navigate the complexity of tracking developmental changes in heterogeneous groups. Outcome measures inform our understanding of phenotypes, development, and decisions concerning the efficacy of treatment approaches; thus, the psychometric properties of these measures must be evaluated. In this study, we provide a preliminary investigation of the psychometric properties of WPCSs derived from naturalistic play samples for preschool-aged boys with autism or FXS. These data provide a first step in establishing the appropriateness of using WPCSs from naturalistic settings in treatment and research studies. Although we specifically considered the use of these metrics in autism or FXS, we believe these metrics will also be useful for children with other NDDs who demonstrate language delays or more variable language trajectories than do children with TD. That said, much like the findings reported in this study, between-group differences in communication development may modify the patterns observed.

Construct Validity

Evaluations of construct validity determine the extent to which variables represent the skill they intend. To establish strong construct validity, one must demonstrate that the outcome variables are significantly correlated with other measures of the same construct (i.e., convergent validity) and not associated, or associated to a lesser extent, with dissimilar constructs (i.e., discriminant validity). For both groups, construct validity was established across all measures considered, suggesting that WPCSs generated from naturalistic play are potentially valid for assessing communication skills in preschoolers with autism or FXS. Moreover, Bland-Altman plots suggested consistency across the two assessment methods. It is important to note that the children included in the present study ranged from having no words/approximations to having mostly three-word utterances or more, but without complex language. Because growth in language can be assessed using different skill types (e.g., gestures, words), depending on the child's communication level, this is a particularly challenging developmental period to assess developmental improvements. Thus, WPCSs are likely to be particularly useful when considering developmental gains within and across these early language stages. That said, additional data are needed to consider the utility of these methods when assessing language skills across a narrower developmental range.

In addition, we were able to replicate prior findings indicating a language strength in boys with FXS, relative to boys with autism, in language skills, when comparisons consider other potential between-group differences (Sterling, 2018; Thurman & Hoyos Alvarez, 2020; Thurman et al., 2017). Indeed, after controlling for effects of age, nonverbal IQ, and overall autism symptom severity, we found that boys with FXS had higher Weighted and Unweighted Communication scores than boys with autism. The Mean Weighted Communication Score did not reach the criterion for a significant effect between groups; it is likely that this effect would have been detected with a larger sample size. Nonetheless, the data from the present study provides initial support that WPCSs in naturalistic samples can be used to assess expected patterns in language development.

Internal Consistency

Evaluations of internal consistency establish the degree of consistency among the items or different parts of a test. In the present study, we first considered internal consistency using split-half reliability methods, with results indicating high reliability for all three WPCSs in both diagnostic groups. In addition, we considered consistency across the four play activities included in the sample. These analyses also indicated strong reliability and correlation scores for both boys with FXS and boys with autism, albeit more variability in the range of scores obtained was observed for boys with autism. Interestingly, for boys with autism, correlations tended to be higher between free play and social interaction activity scores and between requesting and commenting activity scores than when mixing across these pairs. More research is needed to understand whether these findings are related to the social demands of the activity, communication use, and/or influences of the communicative partner.

Play Activity Comparisons

Finally, we considered whether or not WPCSs varied across the different play activities considered. Naturalistic communication samples that consider performance across multiple activities/contexts provide a varied and more representative sample. Nonetheless, it is important to consider the impact these activities/contexts have on communicative performance (Evans & Craig, 1992; Kover, McDuffie, Abbeduto, & Brown, 2012; Miles, Chapman, & Sindberg, 2006). Indeed, activity effects were observed for all WPCSs. Concerning both the Weighted and Unweighted Communication Scores, performance was lower in the commenting activity when compared to all other activities. Mean Weighted Communication score performance during the social interaction activity was lower than performance during all other activities. Together, these findings highlight the effect activity has on both communicative frequency and complexity. Thus, even if scores are strongly correlated across different activities, decisions considering the utility of smaller segments of the sample must be made cautiously, keeping in mind both psychometric appropriateness as well as the impact of differences in the types of communication acts potentially elicited within the segment. For example, for both the Weighed and Unweighted Communication Scores, performance was highest in the requesting activity and lowest in the commenting activity. This pattern of findings is consistent with prior research suggesting that in communication, requests emerge prior to comments (Bruner, 1981; Carpenter, Mastergeorge, & Coggins, 1983); requesting skills are also often targeted in intervention prior to commenting. Thus, patterns of performance are likely change with development and/or intervention experiences.

Limitations and Future Directions

Study analyses provided a preliminary consideration of the psychometric properties of WPCSs in preschool-aged boys with autism or FXS. More extensive investigations are needed, with representative samples, to fully characterize the psychometric utility of WPCSs. For example, future studies assessing the psychometrics of WPCs should also consider test-retest reliability, the extent to which results are consistent with samples collected in other play contexts, and the influence of participant characteristics on results are needed. In addition, it remains unclear the extent to which heterogeneity in language level influences the utility of WPCSs. For example, if everyone in a study sample uses multiple-word utterances, this measure may not be as effective as other measures (e.g., mean length of utterances). Longitudinal data are needed to ensure that WPCSs are sensitive to change over time. Finally, it is important to note that although receptive and expressive language are inter-related, there is a dearth of empirical data validating the use of receptive language measures in individuals with NDD. Investigations considering the utility of receptive language in NDD are also greatly needed.

Conclusions

A lack of outcome measures validated for use in NDDs is a significant barrier for descriptive and treatment research (Berry-Kravis et al., 2018; Kelleher & Wheeler, 2020). We provide a preliminary evaluation of the validity and reliability of weighting procedures to measure communication during naturalistic play in preschool-aged boys with autism or FXS. The present study provides initial support for the psychometric appropriateness of using WPCSs, derived from naturalistic play activities, for preschoolers with autism or FXS. Because the Weighted Communication Scores reflect both frequency and complexity, establishing the psychometrics of all three scores affords investigators multiple options for describing the nature of the language growth over time. Comprehensive investigations of the psychometric properties of WPCSs are needed to determine test-retest reliability, sensitivity to change, and provide insights into the child characteristics that potentially influence psychometric adequacy.

Acknowledgments:

We are indebted to the families for their time, support, and partnership. We also thank the many members of the Laboratory on Language Development and Neurodevelopmental Disabilities who supported the project.

Funding Information:

National Institute of Deafness and Communication Disorders Grant Number: R03DC014543, National Institute of Child Health and Human Development Grant Number: P50HD103526, and National Center for Advancing Translational Sciences Grant Number: UL1TR001860.

References

Abbeduto L, Berry-Kravis EM, Sterling AM, Sherman SL, Edgin JO, McDuffie AS, … Thurman AJ (2020). Expressive language sampling as a source of outcome measures for treatment studies in fragile X syndrome: Feasibility, practice effects, test-retest reliability, and construct validity. Journal of Neurodevelopmental Disorders, 12(1), 10. 10.1186/s11689-020-09313-6 [PubMed: 32204695]

Abbeduto L, McDuffie A, Thurman AJ, & Kover ST (2016). Language Development in Individuals With Intellectual and Developmental Disabilities: From Phenotypes to Treatments. International Review of Research in Developmental Disabilities, 50, 71–118. 10.1016/bs.irrdd.2016.05.006

Adamson LB, Bakeman R, Deckner DF, & Romski M (2009). Joint engagement and the emergence of language in children with autism and down syndrome. Journal of Autism and Developmental Disorders, 39, 84–96. 10.1007/s10803-008-0601-7 [PubMed: 18581223]

American Psychiatric Association. (2013). Diagnostic and statistical manual mental disorders (5th ed.). Washington, DC: American Psychiatric Association.

Bassell GJ, & Warren ST (2008). Fragile X Syndrome: Loss of Local mRNA Regulation Alters Synaptic Development and Function. Neuron, 60(2), 201–214. 10.1016/j.neuron.2008.10.004 [PubMed: 18957214]

Benjamini Y, & Hochberg Y (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple. Journal of the Royal Statistical Society. Series B, 57(1), 289–300.

Berry-Kravis EM, Lindemann L, Jønch AE, Apostol G, Bear MF, Carpenter RL, … Jacquemont S (2018). Drug development for neurodevelopmental disorders: lessons learned from fragile X syndrome. Nature Reviews Drug Discovery, 17(4), 280–299. 10.1038/nrd.2017.221 [PubMed: 29217836]

Bloom L, & Capatides JB (1987). Expression of Affect and the Emergence of Language. Child Development, 58(6), 1513. 10.2307/1130691

Boyd BA, Watson L, Reszka S, Sideris J, Alessandi M, Baranek G, … Belardi K (2018). Efficacy of the ASAP Intervention for Preschoolers with ASD: A Cluster Randomized Controlled Trial. Journal of Autism and Developmental Disorders, 48(9), 3144–3162. 10.1007/S10803-018-3584-Z [PubMed: 29691794]

Brady NC, Fleming K, Romine RS, Holbrook A, Muller K, & Kasari C (2018). Concurrent Validity and Reliability for the Communication Complexity Scale. 27(1), 237–246. 10.1044/2017_AJSLP-17-0106

Braine MDS, & Bowerman M (1976). Children’s First Word Combinations. Monographs of the Society for Research in Child Development, 41(1), 1. 10.2307/1165959

Capone NC, & McGregor KK (2004). Gesture Development. Journal of Speech, Language, and Hearing Research, 47(1), 173–186. 10.1044/1092-4388(2004/015)

Carter AS, Messinger D, Stone W, Celimli S, Nahmias A, & Yoder PJ (2011). A randomized controlled trial of Hanen’s “More Than Words” in toddlers with early autism symptoms. Journal of Child Psychology and Psychiatry, and Allied Disciplines, 52(7), 741–752. 10.1111/J.1469-7610.2011.02395.X [PubMed: 21418212]

Clegg J, Hollis C, Mawhood L, & Rutter M (2005). Developmental language disorders - a follow-up in later adult life. Cognitive, language and psychosocial outcomes. Journal of Child Psychology and Psychiatry, 46(2), 128–149. 10.1111/j.1469-7610.2004.00342.x [PubMed: 15679523]

D’Souza H, & Karmiloff-Smith A (2017). Neurodevelopmental disorders. Wiley Interdisciplinary Reviews: Cognitive Science, 8(1–2), e1398. 10.1002/WCS.1398

Dunn DM, & Dunn LM (2007). Peabody Picture Vocabulary Test, Fourth Edition. Minneapolis, MN: Pearson.

Edwards SL, Rapee RM, Kennedy SJ, & Spence SH (2010). The Assessment of Anxiety Symptoms in Preschool-Aged Children: The Revised Preschool Anxiety Scale. 10.1080/15374411003691701, 39(3), 400–409. 10.1080/15374411003691701

Elliott CD (2007). Differential Abilities Scale - Second Edition (DAS-II). Minneapolis, MN: Psychological Corporation.

Evans JL, & Craig HK (1992). Language Sample Collection and Analysis. 35(2), 343–353. 10.1044/JSHR.3502.343

Fenson L, Marchman VA, Thal DJ, Dale PS, Reznick JS, & Bates E (2007). MacArthur–Bates Communicative Development Inventories: User's guide and technical manual (2nd ed.). Baltimore, MD: Brookes Publishing.

Greenwood CR, Carta JJ, Walker D, Hughes K, & Weathers M (2006). Preliminary Investigations of the Application of the Early Communication Indicator (ECI) for Infants and Toddlers. Journal of Early Intervention, 28(3), 178–196. 10.1177/105381510602800306

Greenwood CR, Walker D, Buzhardt J, Howard WJ, McCune L, & Anderson R (2013). Evidence of a continuum in foundational expressive communication skills. Early Childhood Research Quarterly, 28(3), 540–554. 10.1016/j.ecresq.2013.02.006 [PubMed: 24882940]

Hartley SL, Seltzer MM, Raspa M, Olmstead M, Bishop E, & Bailey DB (2011). Exploring the adult life of men and women with fragile X syndrome: Results from a national survey. American Journal on Intellectual and Developmental Disabilities, Vol. 116, pp. 16–35. 10.1352/1944-7558-116.1.16 [PubMed: 21291308]

Hessl D, Nguyen DV, Green C, Chavez A, Tassone F, Hagerman RJ, … Hall S (2009). A solution to limitations of cognitive testing in children with intellectual disabilities: The case of fragile X syndrome. Journal of Neurodevelopmental Disorders, 1(1), 33–45. 10.1007/s11689-008-9001-8 [PubMed: 19865612]

Homer BD, & Tamis-LeMonda CS (2005). The Development of Social Cognition and Communication. 10.4324/9781315805634

Iverson JM, & Goldin-Meadow S (2005). Gesture Paves the Way for Language Development.

Kelleher BL, & Wheeler AC (2020). Introduction to Special Issue on Outcome Measures for IDD: Where We Have Been, Where We Are Now, and Where We Are Heading. American Journal on Intellectual and Developmental Disabilities, 125(6), 413–417. 10.1352/1944-7558-125.6.413 [PubMed: 33211811]

Kjelgaard MM, & Tager-Flusberg HB (2001). An investigation of language impairment in autism: Implications for genetic subgroups. Language and Cognitive Processes, 16(2–3), 287–308. 10.1080/01690960042000058 [PubMed: 16703115]

Kover ST, McDuffie AS, Abbeduto L, & Brown WT (2012). Effects of sampling context on spontaneous expressive language in males with fragile X syndrome or Down syndrome. Journal of Speech, Language, and Hearing Research, 55(4), 1022–1038. 10.1044/1092-4388(2011/11-0075)

Kuhl PK, & Meltzoff AN (1998). Infant vocalizations in response to speech: Vocal imitation and developmental change. The Journal of the Acoustical Society of America, 100(4), 2425. 10.1121/1.417951

Lord C, Rutter M, DiLavore P, Risi S, Gotham K, & Bishop SL (2012). Autism Diagnostic Observation Schedule-Second edition (ADOS-2).

Maenner MJ, Shaw KA, Baio J, Washington A, Patrick M, DiRienzo M, … Dietz PM (2020). Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. MMWR Surveillance Summaries, 69(4), 1. 10.15585/MMWR.SS6904A1

Mazurek MO, Baker-Ericzén M, & Kanne SM (2019). Brief Report: Calculation and Convergent and Divergent Validity of a New ADOS-2 Expressive Language Score. American Journal on Intellectual and Developmental Disabilities, 124(5), 438–449. 10.1352/1944-7558-124.5.438 [PubMed: 31512950]

McDuffie AS, Thurman AJ, Channell MM, & Abbeduto L (2017). Language disorders in children with intellectual disability of genetic origin. In Handbook of Child Language Disorders: 2nd Edition. 10.4324/9781315283531

McGoey KE, DuPaul GJ, Haley E, & Shelton TL (2007). Parent and Teacher Ratings of Attention-Deficit/Hyperactivity Disorder in Preschool: The ADHD Rating Scale-IV Preschool Version. Journal of Psychopathology and Behavioral Assessment 2007 29:4, 29(4), 269–276. 10.1007/S10862-007-9048-Y

Miles S, Chapman R, & Sindberg H (2006). Sampling Context Affects MLU in the Language of Adolescents With Down Syndrome. 49(2), 325–337. 10.1044/1092-4388(2006/026)

Mullen EM (1995). Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Service.

Rinehart NJ, Cornish KM, & Tonge BJ (2010). Gender differences in neurodevelopmental disorders. In Biological Basis of Sex Differences in Psychopharmacology (pp. 209–229).

Ronald A, & Hoekstra RA (2011). Autism spectrum disorders and autistic traits: A decade of new twin studies. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 156(3), 255–274. 10.1002/AJMG.B.31159

Rosenberg S, & Abbeduto L (1987). Indicators of linguistic competence in the peer group conversational behavior of mildly retarded adults1. Applied Psycholinguistics, 8(1), 19–32. 10.1017/S0142716400000047

Sterling A (2018). Grammar in Boys With Idiopathic Autism Spectrum Disorder and Boys With Fragile X Syndrome Plus Autism Spectrum Disorder. 61(4), 857–869. 10.1044/2017_JSLHR-L-17-0248

Tager-Flusberg HB, Rogers S, Cooper J, Landa R, Lord C, Paul R, … Yoder PJ (2009, June 1). 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, Vol. 52, pp. 643–652. 10.1044/1092-4388(2009/08-0136)

Thurman AJ, & Hoyos Alvarez C (2020). Language Performance in Preschool-Aged Boys with Nonsyndromic Autism Spectrum Disorder or Fragile X Syndrome. Journal of Autism and Developmental Disorders, 50(5). 10.1007/s10803-019-03919-z

Thurman AJ, McDuffie AS, Hagerman RJ, Josol CK, & Abbeduto L (2017). Language skills of males with fragile X syndrome or nonsyndromic autism spectrum disorder. Journal of Autism and Developmental Disorders, 47(3), 728–743. 10.1007/s10803-016-3003-2 [PubMed: 28074353]

Westerveld MF, Gillon GT, & Miller JF (2004). Spoken language samples of New Zealand children in conversation and narration. Advances in Speech Language Pathology, 6(4), 195–208. 10.1080/14417040400010140

Williams KT (2007). Expressive Vocabulary Test (2nd ed.). Minneapolis, MN: Pearson.

Yoder PJ, Stone WL, & Edmunds SR (2020). Parent utilization of ImPACT intervention strategies is a mediator of proximal then distal social communication outcomes in younger siblings of children with ASD: 10.1177/1362361320946883, 25(1), 44–57. 10.1177/1362361320946883

Yoder PJ, Stone WL, Walden T, & Malesa E (2009). Predicting Social Impairment and ASD Diagnosis in Younger Siblings of Children with Autism Spectrum Disorder. Journal of Autism and Developmental Disorders, 39(10), 1381–1391. 10.1007/s10803-009-0753-0 [PubMed: 19449096]

Zimmerman I, Steiner VG, & Pond RE (2011). Preschool Language Scales, Fifth Edition. San Antonio, TX: Pearson.

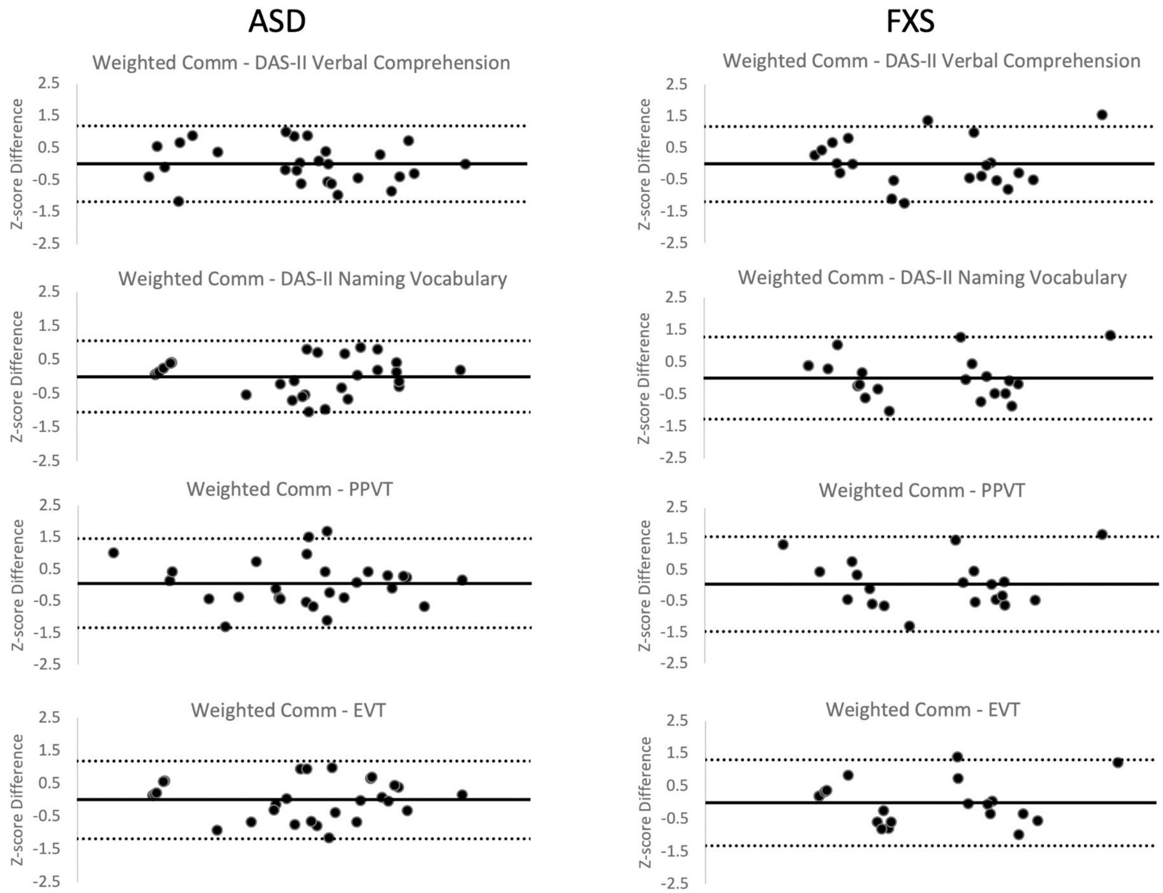

Figure 1.

Weighted Communication Score Bland Altman plots with study validation measures for participants with autism spectrum disorder (ASD; left column) and participants with fragile X syndrome (FXS; right column) Y-axis represents the z-score difference between the two communication methods. X-axis represents the mean z-score across the two communication methods.

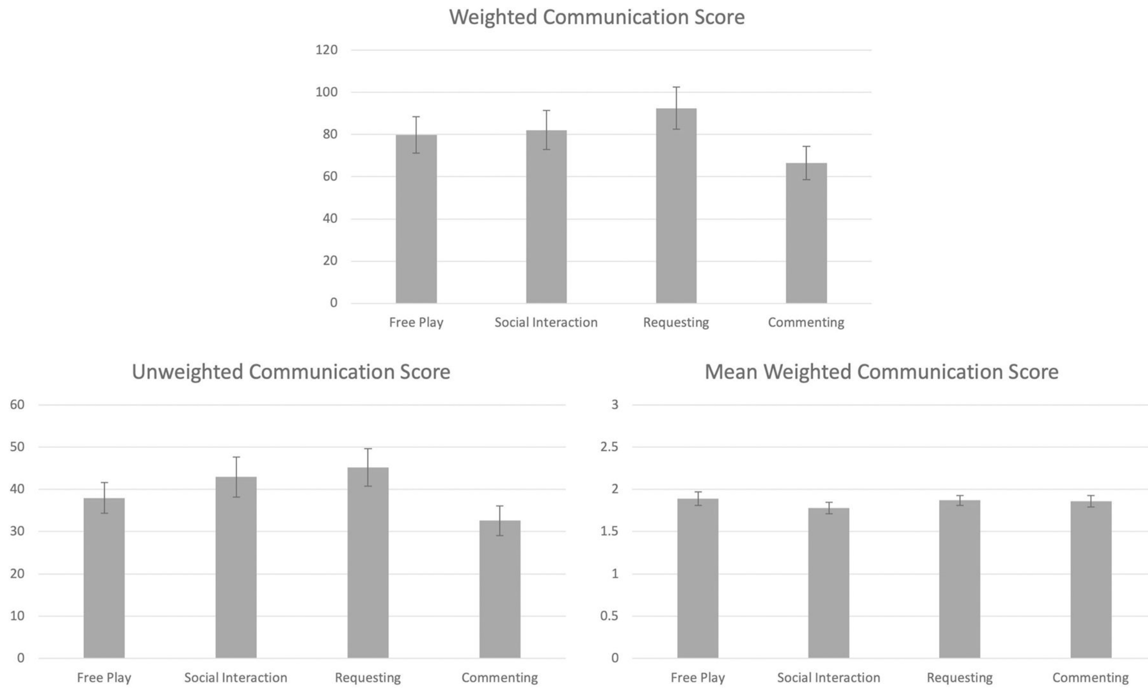

Figure 2.
Main effect of Weighting Procedure Communication Scores as a function of Play Activity

Table 1

Descriptive Statistics as a Function of Diagnostic Group

	Autism (*n* = 28) ***M(SD) or %***	**Fragile X Syndrome (*n* = 21)** ***M(SD) or %***
DAS-II Special Nonverbal Composite SS	75.89 (19.15)	58.57 (16.44)
DAS-II Verbal SS	68.18 (20.36)	62.62 (20.51)
ADOS-2 Comparison Score	7.04 (1.86)	5.79 (2.35)[+]
ADOS-2 Expressive Language Score (minimum criteria)		
(1) No words	10.7%	10.5%
(2) 1 words	10.7%	15.8%
(3) 5 words	14.3%	21.1%
(4) Occasional phrases	14.3%	10.5%
(5) Regular use of 2–3word utterances	21.4%	21.1%
(6) Mostly 3-word utterances (no complex language)	28.6%	21.1%

Note: ADOS-2 = Autism Diagnostic Observation Schedule-2; DAS-II = Differential Ability Scales-II

[+]ADOS-2 missing for 2 participants.

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Ethnic/Racial Composition of Participating Children and Caregivers

	Autism		FXS	
Race	**Child**	**Caregiver**	**Child**	**Caregiver**
Asian	7.1%	14.3%	0%	4.7%
Black/African-American	7.1%	7.1%	4.8%	4.7%
Caucasian	60.7%	60.7%	76.2%	81.2%
Multiracial	21.5%	10.7%	19%	4.7%
Unknown	0%	3.6%	0%	0%
Prefer not to answer	3.6%	3.6%	0%	4.7%
Ethnicity				
Hispanic	25%	7.1%	19%	4.8%
Non-Hispanic	75%	92.9%	81%	95.2%

Table 3

Abbreviated Communication Play Protocol (Adamson et al. 2009) Play Activities

Activity	Toys	Suggestions for parents to facilitate communication
Free Play	Lego set (with blocks, people, animals and vehicles) Farm with accompanying objects School bus with inserting people Inset Puzzle	None provided
Social Interaction (Turn Taking)	Ball Stacking Toy Sorting Toy	Occasionally tease by pausing before giving child a turn
Requesting (Toys out of Reach)	Toy Clock Plush Toy Wooden Frog Pull Toy	Occasionally pretend to misunderstand what child wants
Commenting (Toys in Container)	Action Figure Noise making toy Figure and vehicle Wind-up Toy Pull and go car	Name object and play with it before taking out new object.

Table 4.

Convergent Validity

	Weighted Communication Score		Unweighted Communication Score		Mean Weighted Communication Score	
	ASD	FXS	ASD	FXS	ASD	FXS
DAS-II Verbal Comprehension	.75***	.71***	.69***	.70***	.83***	.77***
DAS-II Naming Vocabulary	.77***	.75***	.69***	.72**	.81***	.88***
PPVT-4	.69***	.68**	.65***	.64**	.72***	.76***
EVT-2	.75***	.74***	.68***	.71**	.85***	.84***

*** $p < .001$

** $p < .01$

* p < .05

Note: DAS-II = Differential Ability Scales-II; PPVT-4 = Peabody Picture Vocabulary Test-4; EVT-2 = Expressive Vocabulary Test-2

Table 5.

Linear Regressions Considering Between-Group Differences in Communication Scores

	B (unstandardized)	*SEB*	*β*	*p-value*	$\eta^2_{partial}$
	Weighted Communication Score **(F(4,42) = 9.57, p < .001, R^2_{adj} = .43)**				
Chronological Age	95.17	31.94	.05	.005 ***	.42
Nonverbal IQ	7.37	1.61	.67	<.001 ***	.58
ASD Symptom Severity	−14.22	13.87	−.14	.311	−.16
Diagnostic Group +	154.73	64.85	.35	.01 *	.35
	Unweighted Communication Score **(F(4,42) = 8.53, p < .001, R^2_{adj} = .40)**				
Chronological Age	36.63	14.21	.31	.01 *	.37
Nonverbal IQ	2.97	0.71	.62	<.001	.54
ASD Symptom Severity	−7.97	6.17	−.18	.203	−.20
Diagnostic Group +	68.17	28.84	.35	.01 *	.34
	Mean Weighted Communication Score **(F(4,42) = 7.57, p < .001, R^2_{adj} = .36)**				
Chronological Age	.17	.07	.31	.01 ***	.37
Nonverbal IQ	.01	.003	.64	<.001 ***	.55
ASD Symptom Severity	−.02	.03	−.09	.550	−.09
Diagnostic Group +	.14	.14	.17	.155	.17

+One-tailed tests

*** p < .001

** p < .01

* p < .05

Table 6.

Correlations Across Play Activities Between Weighting Procedure Communication Scores

	Social Interaction		Requesting		Commenting	
	ASD	**FXS**	**ASD**	**FXS**	**ASD**	**FXS**
			Weighted Communication Score			
Free Play	0.81***	0.83***	0.63***	0.80***	0.62***	0.85***
Social Interaction			0.55***	0.75***	0.65***	0.66**
Requesting					0.73***	0.78***
			Unweighted Communication Score			
Free Play	0.74***	0.82***	0.56**	0.77***	0.55**	0.81***
Social Interaction			0.41*	0.75***	0.63***	0.56**
Requesting					0.72***	0.75***
			Mean Weighted Communication Score			
Free Play	.80***	.90***	.86***	.85***	.77***	.83***
Social Interaction			.76***	.87***	.77***	.93***
Requesting					.74***	.84***

*** $p < .001$

** $p < .01$

* p < .05