Published in final edited form as:

J Dev Behav Pediatr. 2013 May ; 34(4): 245–251. doi:10.1097/DBP.0b013e31828742fc.

Development of an Expressive Language Sampling Procedure in Fragile X Syndrome: A Pilot Study

Elizabeth Berry-Kravis, MD, PhD¹, Emily Doll, BS², Audra Sterling, PhD^{3,5}, Sara T Kover, PhD⁵, Susen M Schroeder, MS³, Shaguna Mathur, MD⁴, and Leonard Abbeduto, PhD⁶

¹Department of Pediatrics, Neurological Sciences, and Biochemistry, Rush University Medical Center, 1725 West Harrison, Suite 718, Chicago, IL 60612, USA

²School of Medicine, Rush University Medical Center, 1653 West Congress Pkwy, Chicago, IL 60612, USA

³Waisman Center, University of Wisconsin-Madison, Madison, WI, USA

⁴Department of Pediatrics, Rush University Medical Center, 1725 West Harrison, Suite 718, Chicago, IL 60612, USA

⁵Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, WI, USA

⁶MIND Institute, University of California, Davis School of Medicine, Sacramento, CA, USA

Abstract

Objective—There is a great need for valid outcome measures of functional improvement for impending clinical trials of targeted interventions for Fragile X syndrome (FXS). Families often report conversational language improvement during clinical treatment, but no validated measures exist to quantify this outcome. This small-scale study sought to determine the feasibility, reproducibility, and clinical validity of highly structured expressive language sampling as an outcome measure reflecting language ability.

Methods—Narrative and conversation tasks were administered to 36 verbal participants (25 males, 11 females) with FXS (age 5–36, mean 18±7). Alternate versions were used with randomized task order, at 2–3 week (mean 19.6±6.4 days) intervals. Audio recordings of sessions were transcribed and analyzed. Dependent measures reflected talkativeness (total number of utterances), utterance planning (proportion of communication (C) units with mazes), articulation (proportion of unintelligible/partly unintelligible C-units), vocabulary (number of different word roots) and syntactic ability (mean length of utterance (MLU) in words). Reproducibility of measures was evaluated with intra-class correlation coefficients (ICC).

Results—All participants could complete the tasks. Coded data were highly reproducible with Pearson correlations $p < 0.01$ for all measures, and ICC values of 0.911–0.966 (conversation) and 0.728–0.940 (narration). Some measures including MLU and different word roots were correlated with expressive language subscale scores from the Vineland Adaptive Behavior Scale (VABS).

Address correspondence to: Elizabeth Berry-Kravis, MD, PhD, Rush University Medical Center, 1725 West Harrison, Suite 718, Chicago, IL 60612, Phone: 312-942-4036, Fax: 312-942-4168, Elizabeth_m_berry-kravis@rush.edu.

Conflicts of Interest and Source of Funding: This work was supported in part by summer student funding from the National Fragile X Foundation and the FRAXA Research Foundation, a Rush University Summer Student Deans Fellowship to ED, and by NIH grant R01HD024356 awarded to L. Abbeduto. For the remaining authors, no conflicts of interest or sources of funding were declared. The authors would like to thank Victor Kaytser for assistance with patient scheduling and visits.

Conclusions—These expressive language sampling tasks appear to be feasible, reproducible, and clinically valid and should be further validated in a larger cohort, as a promising means of assessing functional expressive language outcomes during clinical trials in FXS.

Keywords

fragile X syndrome; *FMR1*; expressive language; outcome measure; autistic disorder

INTRODUCTION

Fragile X syndrome (FXS) is the most common inherited cause of intellectual disability (ID) and the most common known cause of autism, with an estimated prevalence of about 1/4000 (1). FXS results from a trinucleotide repeat (CGG) expansion mutation of greater than 200 repeats (full mutation) in the promoter region of *FMR1* (*Fragile X Mental Retardation 1* gene), which leads to transcriptional silencing and loss or significant reduction of expression of the gene product, FMRP (fragile X mental retardation protein; 2). Loss of FMRP, an RNA binding protein that is a negative modulator of group I mGluR (metabotropic glutamate receptor) and other receptor-activated dendritic translation, results in abnormal brain development, including aberrant dendritic arborization and synaptic plasticity (3–5). Based on recent advances in understanding of the neurobiology of FXS, several pharmacological trials designed to target underlying CNS mechanisms of disease in humans with FXS have been initiated (6–10). Evaluation of such therapeutics, however, is made more difficult by a paucity of objective and well-validated outcome measures of core FXS phenotypes to utilize for measuring efficacy (11). In several trials of targeted treatments, expressive language has been noted by families as an area of improvement (6, 10), however no sufficient measure exists currently to detect such changes. Thus, the main purpose of this study was to develop an expressive language measure that can be used to evaluate the efficacy of interventions in FXS cohorts.

Language impairments are common in FXS, with virtually all aspects of language delayed relative to chronological age expectations in males, although most affected females have significantly less cognitive and language impairment than males, and may have normal language, based on expression of the normal *FMR1* gene from the unaffected X chromosome in a fraction of cells (12). Relative to typically developing and developmentally delayed mental age-matched controls with other disabilities, individuals with FXS show numerous deficits in expressive language, including intelligibility and syntactic complexity, as well as more frequent perseverative language (13–15). Moreover, the profile of expressive language impairments in FXS differs from that in Down syndrome (13–15), suggesting that the profile may be syndrome specific. Hence, there is a profile of expressive language impairments that manifests in FXS due, at least in part (directly or indirectly), to the impact of FMRP deficiency on brain function, and that represents a core FXS phenotype that could be assayed during therapeutic interventions targeting the underlying CNS molecular pathology or core clinical manifestations of FXS (16).

Indeed, an observational measure of expressive language was judged to be a high priority need at a series of NIH meetings (Outcome Measures for Children with Fragile X Syndrome – Parts I and II, May 2008, November 2009), convened to bring together a panel of experts to address the difficult issue of outcome measures for clinical trials for FXS. In fact, procedures for collecting and analyzing expressive language samples have a long history in child language research and for identifying children with language impairments (17). Moreover, these procedures yield dependent measures (e.g., mean length of utterance, or MLU) with excellent psychometric properties (e.g., test-retest, internal consistency, and construct validity) in typically and atypically developing populations 18–19). These

procedures have also been used as outcome measures in behavioral treatments of a variety of populations (20). Moreover, the importance of validating expressive language sampling procedures for FXS is supported by anecdotal reports from families reporting improvements in conversational language during pilot open-label trials, the presence of a core expressive language phenotype in FXS, the clear importance of expressive language to function independently in daily activities in the community and for acquisition of other adaptive skills, and the poor ability of most standardized language tests to measure the kinds of changes in language expected during a relatively short-term intervention in a clinical trial (11,16).

Numerous standardized language tests are available to clinicians and researchers, including those that focus narrowly on a single language domain, such as receptive vocabulary (e.g., Peabody Picture Vocabulary Test-4, 21), and those that are more comprehensive, assessing multiple language domains and performance modalities (e.g., Clinical Evaluation of Language Fundamentals-4; 22). Such tests offer a relatively quick evaluation of performance relative to age expectations, but are often of limited utility for assessing individuals with ID because they are prone to floor effects (23); may be unduly influenced by the co-morbid problems of individuals with ID (e.g., anxiety) thereby obscuring true language skills (16); often have questionable generalizability to everyday language activities (11); and, in many cases, yield summary scores that aggregate performance across a range of language skills, resulting in relative insensitivity to small, but perhaps clinically meaningful, improvements in more narrowly defined domains of language likely to be promoted by targeted treatments (15). Moreover, many standardized language tests evaluate a relatively narrow range of language skills or age ranges, leading to the undesired need to use multiple tests or versions in a single clinical trial population of individuals with ID. For example, in the first clinical trial of a targeted treatment in FXS, which enrolled both males and females with FXS, three different standardized tests directed at different “typical” age ranges were needed to cover the range of language function in the individuals enrolled (24).

Expressive language sampling offers an attractive alternative to standardized language tests. In this procedure, expressive language samples are collected in highly structured, yet naturalistic, interactions with a clinician. These samples are transcribed into electronic text files according to standard conventions and then quickly analyzed using a variety of computer-based algorithms to derive clinical endpoints reflecting a number of dimensions of language skill or atypical language behavior (15). Expressive language samples can be collected quickly and with minimal training of examiners, although transcription can be a time-consuming and resource-heavy process (11). Expressive language sampling has several potential advantages as an assessment tool compared to typical standardized tests. First, the former is more closely aligned with performance in real-world functional contexts (17). Second, expressive language sampling measures are better than standardized tests in terms of discriminating typically developing children from clinically identified children with specific language impairments (SLI; 25). Third, expressive language sampling has been found to be better than standardized tests at identifying language impairments for ethnic and racial minorities, including African-American children (26). Fourth, expressive language sampling can be performed on a cohort with a wide range of language ability without need to use multiple tests. Fifth, the reliability and validity of several dependent measures derived from expressive language samples have been well documented for children with a variety of language impairments (18–19), although these efforts have not extended to individuals with intellectual disabilities. In the present study, we took a first step toward evaluating the feasibility, reproducibility, and clinical validity of expressive language sampling utilizing a structured narrative and conversation task as a source of potential outcome measures for studies of treatment efficacy in individuals with FXS. Procedures for collecting the

expressive language samples were highly scripted to ensure comparability across individuals and times of assessment (18).

METHODS

Participants were recruited from the Fragile X Clinic and Research Program at Rush University Medical Center. All participants were verbal according to caregiver report or clinician experience and had a diagnosis of FXS with molecular confirmation based on DNA analysis. Participants were not allowed to change medications or therapies during the study. The study was approved by the Institutional Review Boards of the participating universities and all participants or their guardians signed informed consent to participate. Thirty six participants with FXS (25 M, 11 F) were recruited and completed test and retest. The age range of participants was 18 ± 7 years (range 5–35; 5–11y=9, 12–17y=9, >18y=18).

Expressive language samples were collected from each participant in two tasks, each a dyadic interaction with an examiner: conversation and narration. The examiner for all participants was a single female research assistant, trained to administer the conversation and narration procedures to ensure comparability across participants. Training consisted of the examiner familiarizing herself with written instructions for administration, viewing of several video-recorded gold-standard administrations, administration to two typically developing children with immediate feedback provided by expert examiners, and then practice administrations to individuals with FXS (not included in the data presented here) with feedback provided by experienced examiners. The entire training procedure was spread over a period of two to three weeks and was no more extensive than required for learning many standardized tests. We have followed these same procedures for training dozens of examiners with varying levels of experience in working with children or individuals with intellectual disabilities. The order of administration of conversation and narration varied randomly across participants. The tasks began after a 5-minute warm-up time in the room interacting with the examiner. Materials for administration of the language sampling procedures and for training examiners are available from the authors upon request.

In conversation, each participant talked with the examiner in an interview-style interaction. The examiner encouraged the participant to talk while trying to minimize her own talk. The examiner followed a script that encouraged talk about a predetermined set of topics and minimized the use of “yes-no” questions in favor of more open-ended prompts (e.g., “tell me about that”). The examiner relied on a standard set of topics, a script for introducing and following up on topics, and a constant order of topic introduction and follow-up questions, thereby ensuring comparability across participants. Switches from one topic to the next, however, were determined by the participant’s level of interest rather than occurring according to a set schedule. The topics included school, teachers, pets, etc. The first topic was always one thought to be of personal interest to the participant based on parent or caregiver report. Follow-up probes were also broad (e.g., “Tell me what you like about your pet,”). Slightly different sets of topics were used for participants who were in school and for those who were no longer in a school setting to ensure that the topics were personally meaningful. Two sets of topics (“A” and “B” versions) were created for each of these groups, so that each participant received a different set at test and retest. Order of administration of the two sets of topics was random across participants (20 of 36 participants got the “A” version first). The conversation was designed to last for 10 minutes, although 4 conversation sessions at test and 5 at retest ended early because of examiner error, scheduling issues, or participant interest. The mean duration (and range) of conversation at test and retest was 9.43 minutes (8.13–10.0) and 9.46 (8.30–10.0), respectively.

In narration, the participant told the story depicted in a wordless picture book, either *Frog Goes to Dinner* (27) or *Frog on his Own* (28). The book used was selected randomly for each participant (20 of 36 participants got the *Frog Goes to Dinner* story first). The participant was first told that he or she would look at the book all the way through one time and then tell the story. In familiarizing the participant with the book, the examiner turned the pages of the book one page at a time, allowing the participant to look at each for about 10 seconds. The examiner then asked the participant to tell the story page by page. The examiner turned from one page to the next five seconds after the participant had finished narrating a page. The examiner used scripted prompts such as, “What about the boy? What’s he doing, thinking, and feeling?”, if the participant failed to narrate the first page. On subsequent pages, the examiner’s prompting was limited to, “What’s happening in this part of the story?”. The task was untimed but generally lasted for 10 to 15 minutes (initial silent viewing and retelling combined). The mean duration (and range) of the retelling portion of the narration at test and retest was 6.45 minutes (3.45–13.16) and 6.40 (3.54–12.31), respectively. Each participant received a different book at test and retest.

The test-retest administrations were separated by two to four weeks. At both test and retest, the participant typically completed the conversation and narration in a single testing session. All expressive language samples were audio-recorded for later transcription and analysis. The VABS was administered to the parent/caregiver during the re-test session or shortly thereafter on another visit. The test-retest interval was 19.6 ± 6.4 days (range 12–28 days for 34 subjects, 37 days for 2 subjects due to vacations).

The language samples were transcribed by experienced transcribers using Systematic Analysis of Language Transcripts software (SALT; 29), a computer program that performs predetermined and customized analyses of text files prepared according to conventions commonly used in child language research (e.g., conventions for segmenting “cats” into root and plural morphemes: “CAT/S”). Each language sample was transcribed first by the “primary” transcriber. A “secondary” transcriber then compared the resulting transcript against the audio-recording and noted any perceived discrepancies on the transcript. The primary transcriber then reviewed the discrepancies and updated the transcript as he/she felt appropriate, thereby creating the final transcript for analysis. The first 10 minutes of each conversation were transcribed, whereas each narrative was transcribed from start to finish regardless of its duration.

All transcribers were trained to high levels of agreement with a gold standard transcriber in advance of this study using an iterative process in which discrepancies with the gold standard were identified and discussed. Training included language samples from individuals of a wide age range and both typically developing children and those with intellectual disabilities and transcribers “passed” only after consistently reaching expected levels of accuracy. This process of training and use of primary and secondary transcribers for each transcript ensures a high degree of accuracy. In fact, agreement between transcripts prepared by different transcribers in our laboratory is typically near 90% for language samples from individuals with intellectual disabilities, including those with FXS (15). Transcribers included in each transcript only speech that had been produced, refraining from simply “assuming” that sounds, such as the plural “s” or the past tense “ed,” had occurred unless they were certain they heard them, and rendering any errors in pronunciation, word selection, or grammatical formation exactly as they occurred. Transcribers segmented all participant talk into C-units, which is defined as an independent clause and any of its modifiers, which can include dependent clauses. Segmentation of talk into C-units rather than utterances has the advantage of more objective criteria and avoids overestimating language ability for long utterances created by simply stringing together simple sentences with the coordinating conjunction “and”. Transcribers also followed standard SALT

conventions for marking mazes (i.e., dysfluencies), such as false starts, repetitions, and filled pauses (e.g., “um” and “er”), as well as any unintelligible portions of speech. They also identified the speaker (subject or examiner) of each C-unit.

Dependent measures were computed using the SALT preset algorithms that automatically create summaries selected by users for each text file containing a transcript (31). The dependent measures of interest in this study reflected amount of talk (total number of C-units), utterance planning/fluency (proportion of complete and intelligible C-units with mazes), intelligibility (proportion of C-units that were either fully or partly unintelligible to the transcriber), vocabulary (number of different word roots in 50 C-units), and syntactic ability (mean length of C-units in words). Additional details regarding the dependent measures can be found in Kover et al 2012 (15).

Reproducibility of measures between sessions was evaluated with interclass correlation coefficients (ICC). Practice effects were assessed by compared means for test and retest with paired-sample *t* tests. Clinical validity was evaluated by determining strength of correlation of measures from the expressive language sample with expressive language scores on the Vineland Adaptive Behavior Scale (VABS). All significance levels reported throughout are for two-tail tests. Although multiple correlations were computed, significance for individual correlations was set at $p < 0.01$ for this pilot exploratory study with a relatively small sample size.

RESULTS

Primary analyses

One participant did not return for retesting and a second participant failed to complete the narration procedures at both test and retest. The tasks were feasible for all other participants, meaning that the subject could complete the tasks over the pre-specified amounts of time, with usable audiotaped data for coding.

The dependent variables from the expressive language samples were highly reproducible, with Pearson correlations significant at $p < 0.01$ for all measures, and with ICCs > 0.7 and significant at $p < 0.01$ for all measures from conversation (Table 1) and narration (Table 2) tasks. The ICCs for MLU were the highest at .960 and .940 for conversation and narration, respectively.

Practice effects from test to retest were minimal. In particular, the differences between the means from test and retest were not significant at $p < .05$, with the exception of mazes in conversation, which nonetheless decreased only from .24 to .22 over the two administrations. The majority of the other *t* tests did not even approach significance (i.e., $p < .15$).

VABS data were obtained for 31 participants. We computed correlations between the dependent measures from the conversation and narration samples and raw scores on the expressive language, receptive language, and written language subscales of the VABS (see Table 3 and 4). MLU in words (in both narrative and conversation tasks) and number of different word roots (narrative task) were correlated with VABS language scores. Proportions of mazes and unintelligible C-units were less strongly correlated with VABS language scores, with fewer significant correlations. The measure of talkativeness – total number of C-units – was not correlated with the VABS scores.

Exploratory analyses

Analyses were also conducted to determine whether the findings for the language sampling procedures varied with subject age and gender. The sample was divided into two age groups (i.e., 17 or younger and 18 or older, $n = 18$ per group) and the analyses reported in the foregoing section were repeated separately for each group. The analyses reported in the foregoing section were also repeated with only the male subjects. In general, few differences emerged relative to the primary analyses. There was a trend, however, toward a reduced magnitude of the test-retest correlations and ICCs for the younger age group for talkativeness and mazes in narration. These analyses, however, should be viewed as exploratory because of the sample sizes and post hoc nature. The analyses are available from the authors.

DISCUSSION

In this small-scale study, expressive language sampling was feasible for verbal individuals with FXS, and measures derived from the coded language samples were highly reproducible and subject to minimal practice effects. The dependent variable means and standard deviations derived from expressive language samples collected in conversation and narration were similar for the first and second administration and ICC values showed strong correlation between test and retest administrations.. Intraclass correlations between test and retest were higher for conversation than narration for most measures, which may have been due to the fact that conversation yielded a larger number of C-units and thus, more data from which to estimate expressive language abilities. At the same time, however, narration appeared impervious to practice effects. Although each sampling procedure is likely to be adequate for a clinical trial, combining the two might yield the most psychometrically sound measure.

The language sampling procedure appears to be feasible within the context of a clinical trial. The conversation and narrative procedures together require less than 30 minutes for administration. The procedures were administered by a research assistant with no previous background in child language research and she required only minimal training from examiners experienced with both the language sampling procedures and the population of interest to achieve fidelity. In addition, most participants could be kept engaged during the language sampling tasks with the structured prompting method. Frustration, transitional anxiety and task avoidance (common problems when testing individuals with FXS) were minimized because a warm-up activity preceded the language samples, the task demands are not high, and the task by its very nature is not associated with a perception of being difficult or not knowing the answer. Although these features minimize testing anxiety, anxiety and behavior might nonetheless interfere with testing during language sampling and could diminish the amount and quality of the language recorded, particularly if the subject has selective mutism. In this pilot study, however, subjects did not change medications or other interventions between the test and retest administrations so as to minimize the impact of variability in the anxiety level on comparisons between results from the two testing sessions as much as possible. The language sampling measures showed a large range with considerable variability across participants suggesting potential for sensitivity to change. The measures also were not subject to ceiling or floor effects over the wide age and ability range of our participants. In particular, all participants were able to produce at least some talk about most conversational topics and pages of the narrative book, produce at least some multi-word C-units, and use at least a small set of words, while producing at least some unintelligible and intelligible C-units and some dysfluent and fluent C-units. At the same time, there is in principle no limit on the amount or type talk possible in these language sampling tasks. Moreover, the mean values in Tables 1 and 2 are not as advanced as found for most typically developing school-age populations (29). Many standardized language

tests, in contrast, are plagued by floor effects for many individuals with intellectual disabilities. Note, however, nonverbal individuals were not recruited into the present study, and the measures would not be appropriate for such participants and could not be used with the subgroup of nonverbal individuals recruited into clinical trials.

The expressive language sampling procedures yielded measures that were correlated with VABS language scores, with MLU in words and number of different word roots showing especially strong correlations. Thus, there is evidence of clinical validity of the language sampling measures in that the measures that would be expected to be most reflective of meaningful language expression (MLU and word roots) were related to functional language levels reported by caregivers both with respect to oral and written language. Because the language sampling task is not measuring receptive language and receptive language is often better developed than expressive language in individuals with FXS, it is not surprising that receptive language scores were less likely to be correlated with the language sampling measures. Talkativeness, as indexed by number of C-units produced, was not correlated with the VABS language scores. This is not surprising because the VABS focuses largely on the maturity of the target individual's communications rather than the sheer number of attempts at communication. Nevertheless, talkativeness is a clinically important variable because individuals who talk more create more opportunities for learning language. Future studies should, therefore, include measures for validation that, unlike the VABS, assess both amount and quality of expressive language. Similarly, intelligibility and mazes were inconsistently related to the VABS; again, possibly reflecting the limited scope of the latter measure. Validation against additional measures of language ability, particularly direct assessment measures rather than informant report measures, would be desirable for further validation of the expressive language sampling procedures (30).

Limitations of the study include a small sample size. This analysis should be repeated in a larger group of sufficient size to analyze gender, age, and ability level effects on measure characteristics and reproducibility. In this regard, exploratory analyses suggested a particular need to examine more closely the psychometric properties of narration as a function of age, a task which we are now undertaking. Further, only one examiner administered the expressive language sampling procedures, which limits external validity, although the procedures used to train that examiner were the same ones we have used to train dozens of other examiners in other studies (13–15). Nevertheless, future work extending the results of this pilot study should include multiple examiners and assess the resulting samples as a function of examiner. Effects of autism status on the characteristics and reproducibility of the expressive language measure were also beyond the scope of this study, but should be investigated. Perseverative language is one of the most characteristic and frequent features of expressive language in both males and females with FXS. Our analysis did not include a measure of perseverative language; however, it is expected that a method for quantification of perseverative language will be developed and applied in future analyses to help validate the method.

In conclusion, data from the present small-scale pilot study suggest that structured expressive language sampling shows evidence of feasibility, reproducibility, and clinical validity in a cohort with FXS, and shows significant promise as an outcome measure for clinical trials to demonstrate improved functional language in association with an intervention. Improved functional language is an endpoint that would clearly reflect a meaningful improvement likely to impact quality of life for individuals with FXS. Given the importance of this endpoint, the versatility of the expressive language task, and lack of ceiling and floor effects, it is anticipated that future research will demonstrate that this assessment can be generalized for use as an outcome measure for interventions in populations with other intellectual disabilities and with autism spectrum disorders.

Acknowledgments

The authors would like to thank Victor Kaytser for assistance with patient scheduling and visits. This work was supported in part by a Rush University Summer Student Deans Fellowship to ED, and summer student funding from the National Fragile X Foundation and the FRAXA Research Foundation and by NIH grant R01HD024356 awarded to L. Abbeduto. For the remaining authors, no conflicts of interest or sources of funding were declared.

References

1. Turner G, Webb T, Wake S, et al. Prevalence of fragile X syndrome. *American Journal of Medical Genetics*. 1996; 64:196–197. [PubMed: 8826475]
2. Oostra BA, Willemsen R. FMR1: a gene with three faces. *Biochimica et Biophysica Acta*. 2009; 1790:467–477. [PubMed: 19233246]
3. D'Hulst C, Kooy RF. Fragile X syndrome: from molecular genetics to therapy. *Journal of Medical Genetics*. 2009; 46:577–584. [PubMed: 19724010]
4. Gross C, Berry-Kravis EM, Bassell GJ. Therapeutic strategies in fragile X syndrome: dysregulated mGluR signaling and beyond. *Neuropsychopharmacology*. 2011; 37:178–195. [PubMed: 21796106]
5. De Rubeis S, Fernandez E, Buzzi A, et al. Molecular and cellular aspects of mental retardation in the Fragile X syndrome: from gene mutation/s to spine dysmorphogenesis. *Advances in Experimental and Medical Biology*. 2012; 970:517–551.
6. Berry-Kravis E, Sumis A, Hervey C, et al. Open-label treatment trial of lithium to target the underlying defect in fragile X syndrome. *Journal of Developmental and Behavioral Pediatrics*. 2008; 29:293–302. [PubMed: 18698192]
7. Berry-Kravis E, Hessel D, Coffey S, et al. A pilot open label, single dose trial of fenobam in adults with fragile X syndrome. *Journal of Medical Genetics*. 2009; 46:266–271. [PubMed: 19126569]
8. Jacquemont S, Curie A, des Portes V, et al. Epigenetic modification of the FMR1 gene in fragile X patients leads to a differential response to the mGluR5 antagonist AFQ056. *Science Translational Medicine*. 2011; 3:64ra1.
9. Berry-Kravis E, Knox A, Hervey C. Targeted treatments for fragile X syndrome. *Journal of Neurodevelopmental Disorders*. 2011; 3:193–210. [PubMed: 21484200]
10. Berry-Kravis E, Hessel D, Rathmell B, et al. Effects of STX209 (arbaclofen) on neurobehavioral function in children and adults with fragile X syndrome: a randomized, controlled, phase 2 trial. *Science Translational Medicine*. 2012 in press.
11. Abbeduto L, Hessel D, Berry-Kravis E, et al. Outcome Measures for Fragile X Syndrome Clinical Trials: Consensus Statement from the NIH Fragile X Coordinating Group. *Journal of Developmental and Behavioral Pediatrics*. 2012 in review.
12. Abbeduto L, Brady N, Kover S. Language development and fragile X syndrome: Profiles, syndrome specificity, and within-syndrome differences. *Mental Retardation and Developmental Disabilities Research Reviews*. 2007; 13:36–46. [PubMed: 17326110]
13. Finestack LH, Abbeduto L. Expressive language profiles of verbally expressive adolescents and young adults with Down syndrome or fragile X syndrome. *Journal of Speech, Language, and Hearing Research*. 2010; 53(5):1334–1348.
14. Kover ST, Abbeduto L. Expressive language in male adolescents with fragile X syndrome with and without comorbid autism. *Journal of Intellectual Disability Research*. 2010; 54(3):246–265. [PubMed: 20146742]
15. Kover ST, McDuffie A, Abbeduto L, et al. Effects of sampling context on spontaneous expressive language in males with fragile X syndrome or Down syndrome. *Journal of Speech, Language, and Hearing Research*. 2012; 55:1022–1038.
16. Abbeduto, L.; McDuffie, A. Genetic syndromes associated with intellectual disabilities. In: Armstrong, CL.; Morrow, L., editors. *Handbook of medical neuropsychology: Applications of cognitive neuroscience*. New York, New York: Springer; 2010. p. 193–221.
17. Abbeduto, L.; Kover, ST.; McDuffie, A. Studying the language development of children with intellectual disabilities. In: Hoff, E., editor. *Handbook of child language research methods*. Malden, MA: Wiley-Blackwell; 2012. p. 330–346.

18. Heilmann J, Nockerts A, Miller JF. Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools*. 2012a; 41(4):393–404.
19. Rice ML, Redmond SM, Hoffman L. Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research*. 2006; 49(4):793–808.
20. Yoder PJ, Molfese D, Gardner E. Initial mean length of utterance predicts the relative efficacy of two grammatical treatments in preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2011 Aug; 54(4):1170–81.
21. Dunn, L.; Dunn, LM. Peabody Picture Vocabulary Test. 4. Minneapolis, Minnesota: Pearson; 2007.
22. Semel, E.; Wiig, E.; Secord, W. Clinical Evaluation of Language Fundamentals. 4. Saddle River, NJ: Pearson; 2003.
23. Mervis CB, Robinson BF. Designing measures for profiling and genotype/phenotype studies of individuals with genetic syndromes or developmental language disorders. *Applied Psycholinguistics*. 2005; 26(1):41–64.
24. Berry-Kravis E, Krause SE, Block SS, et al. Effect of CX516, an AMPA-modulating compound, on cognition and behavior in fragile X syndrome: a controlled trial. *Journal of Child Adolescent Psychopharmacology*. 2006; 16:525–540.
25. Conti-Ramsden G, Crutchley A, Botting N. The extent to which psychometric tests differentiate subgroups of children with SLI. *Journal of Speech, Language, and Hearing Research*. 1997; 40(4): 765–777.
26. Craig HK, Washington JA. An assessment battery for identifying language impairments in African American children. *Journal of Speech, Language, and Hearing Research*. 2000; 43(2):366–379.
27. Mayer, M. *Frog Goes To Dinner*. New York, New York: Dial Books for Young Readers; 1974.
28. Mayer, M. *Frog On His Own*. New York, New York: Dial Books for Young Readers; 1973.
29. Miller, JF.; Iglesias, A. *Systematic Analysis of Language Transcripts (SALT)*, English & Spanish [Computer software]. Version 9. Madison, Wisconsin: University of Wisconsin –Madison, Waisman Center, Language Analysis Laboratory; 2008.
30. Condouris K, Meyer E, Tager-Flusberg H. The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology*. 2003; 12(3):349–358.10.1044/1058-0360(2003/080) [PubMed: 12971823]

Table 1

Expressive Language Scores and Reproducibility for Conversation

Measure	Mean $\pm$ SD Test	Mean $\pm$ SD Retest	ICC ^a	t test ^b
Total Number of C-units	151.67 $\pm$ 44.16	153.97 $\pm$ 39.89	0.911 **	-1.132
MLU in Words	4.37 $\pm$ 1.60	4.28 $\pm$ 1.61	0.966 **	1.039
Proportion C-Units with Mazes	0.24 $\pm$ 0.12	0.22 $\pm$ 0.12	0.927 **	2.158 *
Proportion Unintelligible or Partly Unintelligible C-units	0.11 $\pm$ 0.15	0.09 $\pm$ 0.12	0.939 **	1.417
Number of Different Words Roots	93.97 $\pm$ 30.85	94.82 $\pm$ 32.98	0.884 **	-.242

n = 35

^aIntraclass correlation coefficients between test and retest scores

^bPaired-sample *t* tests comparing test and retest means

* *p* .05,

** *p* < 0.01

Table 2

Expressive Language Scores and Reproducibility for Narration

Measure	Mean $\pm$ SD Test	Mean $\pm$ SD Retest	ICC ^a	t test ^b
Total Number of C-units	91.63 $\pm$ 45.91	87.68 $\pm$ 41.53	0.728**	.303
MLU in Words	4.46 $\pm$ 2.05	4.56 $\pm$ 2.02	0.940**	-.314
Proportion C-units with Mazes	0.19 $\pm$ 0.13	0.18 $\pm$ 0.12	0.788**	.531
Proportion Unintelligible or Partly Unintelligible C-units	0.09 $\pm$ 0.11	0.10 $\pm$ 0.15	0.776**	-.648
Number of Different Words Roots	87.57 $\pm$ 30.24	88.70 $\pm$ 28.20	0.928**	-.358

n=34

^aIntraclass correlation coefficients between test and retest scores

^bPaired-sample *t* tests comparing test and retest means

* *p* .05,

** *p*<0.01

Table 3

Correlations of Conversation Measures at Test with Language Ability Scores on VABS

Variable	Expressive Raw Score	Receptive Raw Score	Written RawScore
Total Number of C-Units	.273	.061	.023
MLU in Words	.463 **	.297	.508 **
Proportion C- Units with Mazes	.129	.179	.350+
Proportion Unintelligible or Partly Unintelligible C-Units	-.339+	-.339+	-.167
Number of Different Words Roots	.312	.176	.508 *

n = 31 or 30 depending on correlation

**
p<.01,*
p<.05,+
p<.06

Table 4

Correlations of Narration Measures at Test with Language Ability Scores on VABS

Variable	Expressive Raw Score	Receptive Raw Score	Written Raw Score
Total Number of C-Units	.107	-.057	.046
MLU in Words	.562 **	.352+	.446 *
Proportion C- Units with Mazes	.337+	.323	.453 *
Proportion Unintelligible or Partly Unintelligible C-Units	-.167	-.411 *	-.086
Number of Different Words Roots	.780 **	.507 *	.821 **

n = 30 or 29 depending on correlation

**
p<.01,*
p<.05,+
p<.06