**Symposium Title:** Development and Validation of Novel Outcome Measures for Children with Developmental Disabilities

**Chair:** Brian A. Boyd  
**Discussant:** James W. Bodfish

**Overview:** As the field moves towards more personalized interventions, there is increasing recognition that the outcome measures commonly used in clinical trials are not always appropriate for individuals with intellectual and/or developmental disabilities (I/DD). Current outcome measures are not adept at detecting the incremental changes seen in individuals with I/DD, have not been consistently validated with these populations, and often fail to measure functional or clinically-relevant outcomes. This has led advocacy organizations, such as Autism Speaks, to convene working groups to tackle this issue as well as led federal agencies, including NIH, to release RFAs specifically focused on the development and validation of outcome measures for I/DD. The overall purpose of this symposium is to present recent efforts across multiple sites and investigators to develop measures that are sensitive to change and suitable for individuals with I/DD. Papers 1, 2 and 3 describe outcome measures developed for children with ASD to detect change in behavioral inflexibility, anxiety, and social-communication, respectively. Paper 4 describes a language outcome measure for children with fragile X or Down syndrome. In general, the measures vary in their methods (e.g., observation, parent report), targeted clinical group, and outcome of interest.

**Paper 1 of 4**

**Paper Title:** Measuring the Functional Impact of Behavioral Inflexibility for Children with ASD  
**Authors:** Brian A. Boyd¹, James W. Bodfish², Luc Lecavalier³, Clare Harrop¹, Allison Whitten², Jill Pritchett³  
**Affiliations:** ¹UNC-Chapel Hill, ²Vanderbilt University, ³The Ohio State University

**Introduction:** Repetitive inflexible behaviors are among the most common forms of aberrant behavior that occur in the context of neurodevelopmental disorders. Yet, the common demands of everyday life require flexible and adaptive responding, which creates difficulties for inflexible children that can range from mild distress to severe behavioral “meltdowns.” It then becomes essential to develop tools that capture the variety and functional impact of these behaviors across a wide age range and functioning levels. Our multisite team is engaged in an iterative but systematic process (Patrick et al., 2007) to develop and validate the Behavioral Inflexibility Scale (BIS), a parent report measure of repetitive inflexible behaviors for children and adolescents (ages 3–17) with ASD. The ultimate goal is for our measure of behavioral inflexibility to also be sensitive to change over time.

**Methods:** This presentation will describe the process used to develop the BIS, which included reviewing existing measures, obtaining expert feedback (n=9), and conducting focus groups (n=55) and/or cognitive interviews (n=8) with parents and clinicians. In addition, the results of a national, web-based survey will be presented to describe the measure’s psychometrics properties. Families registered with the Interactive Autism Network (IAN) (n=934) were contacted to complete the web-based survey. Parents of children with ASD, ages 3 – 17, completed the BIS twice, approximately two weeks apart to examine its factor structure and test-retest reliability. Parents also completed the Social Communication Questionnaire to confirm the child’s autism diagnosis and provided basic demographic information. Parents were reimbursed for survey completion.

**Results:** Preliminary data indicates high internal consistency (α = .95) and test-retest reliability (r=.91). Exploratory and confirmatory factor analyses revealed a single factor with high factor loadings, with all standardized values between .45 and .85. We also plan to present data related to sex bias, and the measure’s performance for children with or without a co-occurring intellectual disability.

**Discussion:** The BIS was designed to measure the everyday occurrence of behavioral inflexibility, impact of that occurrence on family life, and be sensitive to change over time. An essential step in developing an outcome measure is to establish its psychometric properties. A rigorous process has been undertaken to develop the measure, and the initial psychometric data support its reliability and validity. Next steps will involve examining its potential as a treatment outcome measure.
References:


**Paper 2 of 4**

**Paper Title:** Toward Outcome Measurement of Anxiety in Youth with ASD

**Authors:** Luc Lecavalier\(^1\), Robert T. Schultz\(^2\), Michael Aman\(^1\), Karen Bearss\(^3\), Judith Miller\(^2\), Christopher Taylor\(^1\), John Herrington\(^2\), Michael Edwards\(^1\), and Lawrence Scahill\(^3\)

**Affiliations:** \(^1\)The Ohio State University, \(^2\)Children’s Hospital in Philadelphia, \(^3\)Emory University

**Introduction:** The limited number of valid outcome measures for people with DD is a major impediment for intervention research. Impairing anxiety is common in youth with ASD and a plausible treatment target, but poorly studied. There are well-accepted outcome measures for anxiety in typically developing youth. Given the language and cognitive deficits in ASD, these measures are not suitable for youth with ASD without modification. We set out to develop a parent-rated and a clinician-rated instrument for measuring anxiety in youth with ASD who participate in clinical trials. We used a mixed method approach in line with the FDA monograph on Patient-Reported Outcome Measures (Patrick et al., 2007).

**Methods:** We first conducted six focus groups with 48 parents of 45 children (aged 3–17 years) with ASD and at least mild anxiety. This information was used to help generate items for a parent rating scale and refine an existing clinician-rated instrument (*Pediatric Anxiety Rating Scale* [PARS]). Ratings were then collected on 990 children and adolescents (81% boys, mean age = 11.3 years, SD = 3.4 years) on a preliminary rating scale. These data were used for item analysis and to assess factor structure. We then invited 90 children for in-person assessments. The goals of these assessments were to measure the stability of the rating scale and clinical interview (3 measurements over a 4-week period) and to assess their validity by correlating scores with other demographic and clinical variables. Finally, a high-anxiety subgroup is being compared to a low-anxiety group on physiological measures of heart rate [HR] and heart rate variability [HRV] at rest and under mildly anxiety provoking conditions.

**Results:** Systematic coding of the focus group transcripts identified broad themes reflecting the situations and events that trigger anxiety in children with ASD, the behavioral manifestations of anxiety in youth with ASD, the parent and the child’s response to anxiety, and broad behavioral patterns that could be associated with anxiety. This information helped investigators generate 52 candidate items for the parent-completed rating scale. These items, combined with DSM-based items taken from the *Child and Adolescent Symptom Inventory* (CASI), were submitted to a series of exploratory and confirmatory factor analyses. A number of items were eliminated based on these analyses, resulting in a much shorter scale. Preliminary correlations with measures of repetitive behavior, compulsive behavior, behavior problems, and ASD symptoms suggest good validity. Preliminary analyses also indicate good stability. Data collection on HR and HRV is under way.

**Discussion:** The measurement of anxiety in ASD is fraught with uncertainty. Our mixed method approach at developing and validating instruments yielded short instruments for outcome measurement in treatment studies targeting anxiety. Preliminary evidence supports their reliability and validity.

**Paper 3 of 4**

**Paper Title:** Measuring Changes in Social Communication in Children and Adults with ASD

**Authors:** Catherine Lord\(^1,2\) and Rebecca Grzadzinski\(^1,2\)

**Affiliations:** Center for Autism and the Developing Brain, Weill Cornell Medicine\(^3\), Columbia University\(^2\)

**Introduction:** The field of ASD intervention research lacks outcome measures sensitive enough to capture subtle changes in social communication behaviors over the course of a short-term intervention. Most measures currently used are either reliant on reports of symptoms from parents or therapists, who are generally not blind to most behavioral interventions, or were developed by researchers to measure a specific intervention-targeted behavior, which is appropriate, but not the same as testing
whether the behavior occurs outside of the context in which it was taught. These limitations in outcome measures affect the range of social communication behaviors captured and hinder the ability of researchers to replicate findings and merge samples for more complex analyses. To address these issues, our group developed the Brief Observation of Social Communication Change (BOSCC).

**Methods:** The BOSCC is a 12-minute videotaped interaction between a child and adult in which the adult is given some, but minimal, guidelines for how to lay out standard materials. It is intended to be administered and scored by different people who do not need lengthy training and can remain blind to all aspects of the study. The BOSCC coding scheme was applied to 177 parent-child free-play videos from 56 toddlers and preschoolers with diagnoses of ASD. All children were using phrase speech or less and were participating in various intervention trials active at two sites.

**Data Analysis:** Exploratory Factor Analyses (EFA) were conducted, providing BOSCC domains, as well as intraclass correlations (ICCs) for Test-Retest (TRT) and Inter-rater reliability (IRR). Growth curve models were fit to all available data for each child’s BOSCC and ADOS Calibrated Symptom Severity (CSS) scores. Exploratory analyses were also conducted comparing BOSCC scores in children who did and did not demonstrate changes in Vineland Adaptive Behavior Scales- Communication (VABS) and Mullen Scales of Early Learning Receptive Language (MSEL) scores, two of the most common outcome measures in early intervention studies.

**Results:** EFA results support a two-factor model, consistent with the social communication and RRB domains seen in the ASD symptom literature. ICCs computed for TRT and IRR data ranged from 0.81-0.98 and from 0.94-0.99, respectively. Results of whole sample random effects analyses showed that BOSCC scores changed more rapidly than ADOS CSS scores. Children who demonstrated improvements (increases) on the VABS and MSEL, two common outcome measures in early intervention studies also demonstrated improvements (decreases) in BOSCC scores. In one analysis, the BOSCC was more sensitive to change than CGIs by the examiner. Additional development of a BOSCC scale for older, verbal children is underway and will be described.

**Conclusions:** This work provides initial evidence for the utility of the BOSCC as a treatment response measure. The BOSCC may be one way to address the need for an outcome measure that is sensitive enough to capture subtle changes while also standardized and flexible for use across studies. We hope that the BOSCC can bolster the value of other measures commonly used for treatment response in order to aid in identifying efficacious treatments, monitoring progress, and tailoring individualized treatment plans for children with ASD.

---

**Paper Title:** Expressive Language Sampling as an Outcome Measure for Clinical Trials

**Authors:** L. Abbeduto¹, A.J. Thurman², A. McDuffie¹, E. Berry-Kravis², A. Sterling³, J. Edgin⁴, S. Sherman⁴

**Affiliations:** UC Davis MIND Institute¹, Rush University², University of Wisconsin-Madison³, University of Arizona⁴, Emory University⁵

**Introduction:** Drugs are in development to treat core symptoms of fragile X syndrome (FXS) and Down syndrome (DS). Disorder-specific behavioral interventions are also being developed. Evaluation of the efficacy of pharmaceutical and behavioral interventions, however, is hampered by a lack of psychometrically sound outcome measures (Berry-Kravis et al., 2013). In the current study, we present preliminary data from a large multi-site project evaluating the adequacy of expressive language sampling (ELS) as an outcome measure. Standardized ELS procedures have been developed for two tasks: (a) conversation with an examiner and (b) narration of a wordless picture book (Abbeduto et al., 2012). The psychometric properties of ELS, however, need to be documented.

**Research Objectives:** (1) Examine potential practice effects for ELS, (2) Examine the test-retest reliability of ELS, and (3) Examine the construct validity of ELS.

**Method:** Data have been collected from 200+ participants, ages 6 to 23 years, across five sites. Analyses are ongoing, but will be completed by the Conference. The findings reported in this abstract are for a subset of the FXS (n=25, Mean age = 12.8 years) and DS participants (n=20; Mean age = 13.8 years). Participants are native English speakers, use speech as their primary mode of communication, produce at least some multiword utterances, and have no more than a mild hearing loss.
Conversational and narrative samples were collected from each participant at two time points, four weeks apart, to examine practice effects and test-retest reliability. Samples were audio-recorded and transcribed using SALT software. Dependent variables were computed from the transcripts.

Raw scores from norm-referenced standardized language tests were used to evaluate the (convergent) construct validity of the ELS variables. Different blinded examiners administered ELS and standardized language tests.

**Results:** Analyses were conducted separately for FXS and DXS and for conversation and narration. In terms of practice effects, scores changed significantly from T1 to T2 only for vocabulary in conversation for DS. In terms of test-retest reliability, scores on T1 and T2 were significantly correlated for all ELS measures in conversation and narration for both FXS and DS. In terms of construct validity, numerous significant correlations emerged between the ELS variables and their corresponding convergent standardized tests for conversation and narration and for FXS and DS, but most consistently for intelligibility, vocabulary, and syntax.

**Discussion:** ELS has several advantages compared to standardized language tests. ELS procedures yield data more reflective of performance in meaningful real-world contexts. Numerous dependent measures, each reflecting a different aspect of spoken language, can be computed from a single language sample. ELS procedures have limited floor effects for individuals producing at least some multiword utterances. Preliminary psychometric data suggest ELS holds promise as an outcome measure for FXS and DS. The next step, which is underway, is to evaluate sensitivity to change of ELS measures and variations in psychometrics as a function of participant characteristics.

**References:**