Evaluation of Expressive Language Sampling-Outcome Measures for Individuals with ASD

Lizbeth H. Finestack, PhD, CCC-SLP
Department of Speech-Language-Hearing Sciences
finestack@umn.edu

Funding & Disclosures

- PI: Leonard Abbeduto, UC Davis, MIND Institute
  - National Institutes of Health (R01HD07436-03S1)
  - Simons Foundation (SFARI 383668)
- Received honorarium and travel expense coverage

Outline

- Clinical trials and outcome measures
- Language as outcome in treatment studies
- Measuring language development
- Expressive Language Sampling project
- Preliminary Findings

Clinical Trials & Outcome Measures
Clinical Trials

- Clinical trials are carefully designed research studies that answer specific questions about the effectiveness of:
  - Preventive techniques
  - Medications
  - Therapies and treatments
  - Surgical techniques

[https://www.health.umn.edu/research/clinical-trials](https://www.health.umn.edu/research/clinical-trials)

Clinical Trial Failures??

- Uncontrolled variability within and across treatments
- Limitations on our understanding of disease mechanisms
- Lack of supporting behavioral or educational interventions
- Lack of psychometrically sound, meaningful, sensitive measures of change in core symptoms

Outcome Measures

- Can be:
  - Generic or disease-specific
  - Physician- or patient-based
  - Direct or indirect
  - Cover all or specific aspects of the condition


Outcome Measures

- Need to:
  - Be valid
  - Be reliable
  - Have ability to detect change
  - Minimize bias
  - Be relevant to key stakeholders:
    - Patients and their families, public, health care professionals, policy makers

Call for Core Outcome Set (COS)
- Development and use of an agreed standardized collection of outcomes which should be measured and reported in all trials for a specific condition

Supported by:
- GRADE (Grading of Recommendations Assessment, Development and Evaluation) Group
- Cochrane Reviews
- World Health Organization (WHO)

Advantages
- Reduce heterogeneity in reported outcomes between trials
- Reduce risk of outcome reporting bias
- Statistical power for meta-analyses increase

Example from Fragile X Syndrome

Outcome Measures for Clinical Trials in Fragile X Syndrome
Elizabeth Berry-Kravis, MD, PhD; David Hecht, PhD;1,2 Leonard Abbeduto, PhD;1,2,3 Allen L. Reiss, MD; Andrea Beckel-Mitchener, PhD; Tina K. Urv, PhD, and Outcome Measures Working Group

The Process

Establish Working Groups of Experts
- Develop Conceptual Framework for each Group
- Identify Available Assessments
- Assess Measurement Properties
- Identify Measures

- Cognitive
- Emotional
- Physical
- Social

- Activities completed through a series of virtual meetings
- Two face-to-face meetings

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Language as an Outcome in Treatment Studies
Pharmaceutical treatments are currently being developed and tested for individuals with intellectual disabilities, including ASD. Language could be a useful outcome measure (and intervention target) because improvements in language are meaningful and functional for the individual and his/her family.

Language in Treatment Studies

Improvements in language are meaningful and functional.
Language may improve because:
- New language skills/elements learned
- Improvements in foundational domains
- Reductions in maladaptive behaviors/mental health symptoms

Limitations of Standardized Tests

- Generalizability to natural social interaction unclear
- Floor effects for testing people with intellectual disabilities
- Narrow age range
- Expressive items/tests
  - Highly scaffolded
  - Often “meta” component

Measuring Language Development
Expressive Language Sampling (ELS)

- Collection of brief, but representative samples of expressive language during social interaction
- Transcribe and analyze for a range of components, skills, patterns of behavior
- History of use in clinical contexts and research describing the development of and impairments in spoken language

Syntax (Kovas et al., 2002)

<table>
<thead>
<tr>
<th></th>
<th>TD</th>
<th>DS</th>
<th>FXS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Length of Utterance (Morphemes)</td>
<td>3.94</td>
<td>3.56</td>
<td>3.46</td>
</tr>
</tbody>
</table>

Vocabulary (Kovas et al., 2002)

<table>
<thead>
<tr>
<th></th>
<th>TD</th>
<th>DS</th>
<th>FXS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Different Word Roots</td>
<td>16.2</td>
<td>11.5</td>
<td>15.2</td>
</tr>
</tbody>
</table>

ELS Contexts

- Freeplay with parent
- Structured play with examiner
- Conversation with peers
- Retell a story
- Picture description
- Communication probes (e.g., ADOS)
ELS Advantages

- Performance is generalizable to real-world contexts
- Noncompliance rates are low
- Numerous dependent measures can be computed from a single sample
- Samples may be brief
- Can be used with children and adults
- Limited floor effects for individuals with complex needs
- Samples may be brief
- Simple to collect

Selecting ELS Contexts

- Representativeness influenced by materials, partner behavior, and genre/task
- Need to “standardize” sampling procedures
  - Avoid constraints that lead to “underestimates”
  - Avoid scaffolding that leads to “overestimates”
- Need to sample in multiple speaking tasks for a comprehensive picture

ELS Outcome Measure

- Important to standardize the procedures for collecting samples of spoken language
  - Constrain content through materials selected
  - Constrain examiner behavior through use of scripts
- Standardization allows consistency across:
  - Examiners
  - Sites
  - Participants
  - Assessment occasions

“Standardized” Procedures

- Topics in conversation; Story in narration
- Examiner behavior highly scripted but still responsive and natural
- Alternate forms
- 10-15 minutes per task

Idiosyncratic topic:
- I was talking to your Mom/Dad/Teacher and they told me that you... [TOPIC]
- That sounds so interesting/like so much fun. Tell me about that.

Idiosyncratic topic: follow-up
- I'd like to hear more about...
  - [TOPIC]
  - what you like about [TOPIC]
  - why you like [TOPIC]
  - how you [TOPIC]
  - what happens when you [TOPIC]

School Day
- So, what grade are you in?
  - It's been a long time since I was in school so I really don't remember very much about the [participant's grade level].
- Tell me what you did in school yesterday [or the last day the participant was in school].
  - Tell me everything you can remember
**Conversational Language Sample**

- **School Day: follow-up**
  - What's your favorite part of school?
    - Tell me all about that.
    - Tell me why you like doing that.
  - What's the first thing you do when you get to school?
    - Tell me about what you do first
  - What's the last thing you do at school before you go home?
  - Is there anything you don’t like about school?
    - Tell me why you don’t like that.

**Narrative Language Sample**

- The goal of the clinician is to give as little input as possible
- 2 passes through the book
  - First pass: no talking just, look at pictures
  - Clinician always turns the pages
Narrative Language Sample

- 2nd pass: Child tells the story
  - 5-7 seconds per page (no matter what!)
- 1st page
  - How does the story start?
  - Hierarchy of prompts

PROMPT A [If no response or IDK]
- What’s happening in this part of the story?

PROMPT B [minimal response]
- That’s a good start. Tell me a little more about what’s happening in this part of the story.

Narrative Language Sample

- PROMPT C [if no or minimal response to prompt A or B]
  - What about the boy? What’s he doing?
  - What about the frog? What’s he doing?
  - What about the turtle? What’s he doing?
  - Anything else? (use sparingly)
  - Okay, here’s the next page.

ELS Outcome Measure

- Evidence of validity and clinical utility
  - Distinguishes DS from FXS and TD (and SLI/DLD from TD)
  - Distinguishes between subgroups within a syndrome
- Labor-intensive in transcription and coding, but increasingly automated
- Adequacy of psychometric properties?
To examine the psychometric properties of measures derived from ELS procedures in FXS, DS, and ASD.

- Practice effects
- Internal consistency
- Test-retest reliability
- Validity
- Sensitivity to change
- Diverging

To evaluate differences in the psychometric properties of ELS as a function of participant characteristics

**Participants**
- 275 6- to 23-year olds with either DS, FXS, or ASD
- Three age groups: 6-11, 12-17, 18-23 years
**Project Design**

- Initial visit
- Retest at 4 weeks (+/- week)
- Follow-up at 1-2 years

**EXPRESSIVE LANGUAGE SAMPLING CONSORTIUM SITES**
(n = 50 per site)

**ELS Contexts**

- **Conversation**
  - Script with topics
  - Duration of sample: 12 mins

- **Narration**
  - Wordless picture books with limited scaffolding
  - Duration of sample: Untimed but ~ 10 mins

- **Autism Diagnostic Observation Schedule, 2nd edition (ADOS)**
  - Series of prompts to elicit behavioral "red flags"
  - Duration of sample: 30-45 mins

**ADOS Language Sample**

[Image of ADOS-2 booklets]
ADOS Language Sample
- Narrative
- Conversation
- Expository Discourse
- Free Play

ADOS: Narrative

ADOS: Narrative

Cartoon
Throughout the ADOS session, examiner is looking for opportunities for conversion.

Jump into topics as they come up by the child.

Also provide opportunities with general questioning.

Examiner "leads":
- "I did something really cool yesterday..."
- "Oh that reminds me of something funny..."

**ADOS: Structured Conversation**

**Description of a Picture:**

- Interview: Relationships, school, emotions

**ADOS: Play**

- Make Believe Play
- Joint Interactive Play

**ADOS: Play**

- Free Play: younger children (does not yet use complex grammar, 5 years or younger)
**ADOs: Expository**

- **Demonstration Task**
  - “Pretend like I don’t know how to brush my teeth, you are going to show and tell me how to do it.”

**Transcription**

- Audio-recorded samples are transcribed into electronic text files
- Transcripts are analyzed to generate clinical endpoints reflecting the dimensions of interest

---

**Primary Dependent Measures**

- **Talkativeness**: number of utterances attempted per minute
- **Intelligibility**: proportion of utterances that are partly or completely unintelligible
- **Dysfluency**: proportion of utterances that contain a verbal dysfluency/false start/filler
- **Lexical Diversity**: number of different words in 50 utterances
- **Syntactic Complexity**: mean length of utterance (morphemes)

**Convergent Validity Measures**

- **Talkativeness**: Vineland Adaptive Behavior Scales: Expressive Communication
- **Intelligibility**: Goldman-Fristoe Test of Articulation
- **Dysfluency**: CELF: Rapid Auditory Naming
- **Lexical Diversity**: CELF: Expressive Vocabulary
- **Syntactic Complexity**: CELF: Formulated Sentences
### Preliminary Findings

#### Conversation and Narrative Contexts

- 73 participants with ASD in database, to date
  - 22% female; 78% male

#### Participants with Conversational Samples at Times 1 & 2

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td>33</td>
<td>12.14</td>
<td>3.52</td>
<td>6.19</td>
<td>18.88</td>
</tr>
<tr>
<td>Nonverbal IQ</td>
<td>41</td>
<td>42</td>
<td>131</td>
<td>89.02</td>
<td>23.73</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>40</td>
<td>32</td>
<td>130</td>
<td>89.38</td>
<td>25.72</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>40</td>
<td>40</td>
<td>132</td>
<td>90.78</td>
<td>23.30</td>
</tr>
<tr>
<td>ADOS Severity</td>
<td>42</td>
<td>1</td>
<td>10</td>
<td>6.43</td>
<td>2.21</td>
</tr>
</tbody>
</table>

#### Participants with Narrative Samples at Times 1 & 2

<table>
<thead>
<tr>
<th>Descriptive</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td>32</td>
<td>12.26</td>
<td>3.28</td>
<td>6.19</td>
<td>18.88</td>
</tr>
<tr>
<td>Nonverbal IQ</td>
<td>40</td>
<td>87.78</td>
<td>25.48</td>
<td>42</td>
<td>131</td>
</tr>
<tr>
<td>Verbal IQ</td>
<td>39</td>
<td>88.03</td>
<td>27.34</td>
<td>32</td>
<td>130</td>
</tr>
<tr>
<td>Full Scale IQ</td>
<td>39</td>
<td>89.49</td>
<td>25.27</td>
<td>40</td>
<td>132</td>
</tr>
<tr>
<td>ADOS Severity</td>
<td>41</td>
<td>6.39</td>
<td>2.18</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>
### ASD Test-Retest Practice Effects

<table>
<thead>
<tr>
<th></th>
<th>Conversation</th>
<th>Narration</th>
<th>P (df)</th>
<th>P (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talkativeness (utt/min)</td>
<td>11.05 (5.04)</td>
<td>10.61 (4.09)</td>
<td>.19 (2.24)</td>
<td>8.85 (3.60)</td>
</tr>
<tr>
<td>Intelligibility (% C &amp; I utt)</td>
<td>3.87 (5.51)</td>
<td>3.66 (5.50)</td>
<td>.60 (4.03)</td>
<td>4.70 (7.94)</td>
</tr>
<tr>
<td>Dysfluency (% utt with maze)</td>
<td>33.18 (14.29)</td>
<td>33.38 (14.32)</td>
<td>.89 (4.06)</td>
<td>33.34 (17.68)</td>
</tr>
<tr>
<td>Lexical Diversity (NDW)</td>
<td>130.42 (45.96)</td>
<td>132.52 (40.60)</td>
<td>.46 (2.10)</td>
<td>108.74 (49.67)</td>
</tr>
<tr>
<td>Syntax (MLU)</td>
<td>6.19 (1.99)</td>
<td>6.37 (2.79)</td>
<td>.17 (1.81)</td>
<td>7.44 (4.46)</td>
</tr>
</tbody>
</table>

Paired-samples T-tests; n = 50-53

### DS Test-Retest Reliability

<table>
<thead>
<tr>
<th></th>
<th>Conversation ICC (p)</th>
<th>Narration ICC (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talkativeness (utt/min)</td>
<td>.89 (.000)</td>
<td>.74 (.000)</td>
</tr>
<tr>
<td>Intelligibility (% C &amp; I utt)</td>
<td>.90 (.000)</td>
<td>.76 (.000)</td>
</tr>
<tr>
<td>Dysfluency (% utt with maze)</td>
<td>.65 (.000)</td>
<td>.50 (.000)</td>
</tr>
<tr>
<td>Lexical Diversity (NDW)</td>
<td>.88 (.000)</td>
<td>.92 (.000)</td>
</tr>
<tr>
<td>Syntax (MLU)</td>
<td>.87 (.000)</td>
<td>.88 (.000)</td>
</tr>
</tbody>
</table>

Intraclass Correlation Coefficients
**Preliminary Data**

- Practice Effects
- Test-Retest Reliability
- Construct Validity

**Convergent Reliability: Conversation**

<table>
<thead>
<tr>
<th></th>
<th>Vineland: EC</th>
<th>GFTA: SIW</th>
<th>CELF: RAN</th>
<th>CELF: EV</th>
<th>CELF: FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talkativeness</td>
<td>-.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intelligibility</td>
<td>-.46*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dysfluency</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical Diversity</td>
<td>.79*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntax</td>
<td>.46*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at $p = .001$ level

**Convergent Reliability: Narrative**

<table>
<thead>
<tr>
<th></th>
<th>Vineland: EC</th>
<th>GFTA: SIW</th>
<th>CELF: RAN</th>
<th>CELF: EV</th>
<th>CELF: FS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talkativeness</td>
<td>-.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intelligibility</td>
<td>-.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dysfluency</td>
<td>-.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical Diversity</td>
<td>.58*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntax</td>
<td>.54*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significant at $p = .001$ level

Controlling for age ($n = 27$)
Significant at $p = .001$ level
### Convergent Reliability: Narrative

<table>
<thead>
<tr>
<th></th>
<th>Vineland</th>
<th>GPTA-SIN</th>
<th>CELF-RAN</th>
<th>CELF-EV</th>
<th>CELF-FB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talkativeness</td>
<td>- .47*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intelligibility</td>
<td></td>
<td>- .58*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dysfluency</td>
<td></td>
<td></td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical Diversity</td>
<td></td>
<td></td>
<td>.46*</td>
<td></td>
<td>.37*</td>
</tr>
<tr>
<td>Syntax</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Controlling for age (n = 27)
Significant at p = .001 level

### Preliminary Findings

- **Minimal practice effects**
- **Excellent test-retest reliability**
- **Evidence of construct validity for some measures**
  - Lexical diversity, syntax, and intelligibility – Conversation
  - Lexical diversity and syntax - Narrative
  - Stronger convergence when control for age

### Immediate Next Steps

- Analyze ADOS psychometric properties
- Analyze psychometrics relative to
  - Person characteristics
  - Sample characteristics (e.g., DS vs. FXS vs. ASD)
- Evaluate long-term outcomes
- Additional outcome measures to be derived, including pragmatic indicators and language composites
- Examine site differences

### Long-term Next Steps

- Develop and evaluate transcription/coding automation
- Examine consistency across languages
- Evaluate implementation
Acknowledgements

University of Minnesota
- Katie Bangert, M.A., CCC-SLP
- Timothy Huang, M.S., CCC-SLP
- Elsa Temte
- Marcelia Grachulska
- Meredith Henke
- Asha Koritala
- Megan Larsen
- Natalie Rosen

ELS Consortium
- Leonard Abbeduto, UC Davis MIND Institute
- Sara Kover, University of Washington
- Elizabeth Berry-Kravis, Rush University Medical Center
- Audra Sterling, University of Wisconsin-Madison
- Stephanie Sherman, Emory University
- Jamie Edgin, University of Arizona
- Lizbeth Finestack, University of Minnesota-Twin Cities

Acknowledgements

Hundreds of Families!!!

Acknowledgements

Thank you!

Lizbeth H. Finestack, PhD, CCC-SLP
University of Minnesota
Department of Speech-Language-Hearing Sciences
finestack@umn.edu/finestackclil.com