

American Journal on Intellectual and Developmental Disabilities

Validation of the Observer-Reported Communication Ability measure for individuals with Angelman syndrome

--Manuscript Draft--

Manuscript Number:	AJIDD-D-21-00023R5
Article Type:	Research Report
Keywords:	Communication, Angelman syndrome, outcome measure, psychometrics, cognitive testing
Corresponding Author:	Christina K Zigler, PhD, MEd Duke University School of Medicine Durham, North Carolina UNITED STATES
First Author:	Christina K Zigler, PhD, MEd
Order of Authors:	Christina K Zigler, PhD, MEd Li Lin, MS Molly McFatrach, MPH Nicole Lucas, BS Kelly L Gordon, MS, CCC-SLP Harrison N Jones, PhD Allyson Berent, DVM, DACVIM Jennifer Panagoulas, RAC Paula Evans Bryce B Reeve, PhD
Manuscript Region of Origin:	UNITED STATES
Abstract:	There is a critical need for high-quality clinical outcome assessments to capture the important aspects of communication ability of individuals with Angelman syndrome (AS). To center the perspective of caregivers, our team developed the novel Observer-Reported Communication Ability (ORCA) measure using best practice guidelines, with the goal of developing a measure that could be administered to caregivers directly without the need for a certified administrator for use in clinical trials. To refine the draft measure, we conducted two rounds of cognitive interviews with 24 caregivers and a quantitative study including 249 caregivers. The results from both studies support the overall content validity, construct validity, and the reliability of the ORCA measure for individuals with AS > 2 years old for use in research contexts. Future work should explore the responsiveness of ORCA measures to changes over time in a diverse sample.

Validation of the Observer-Reported Communication Ability (ORCA) Measure for Individuals with
Angelman Syndrome

**Validation of the Observer-Reported Communication Ability (ORCA) Measure for
Individuals with Angelman Syndrome**

ABSTRACT

There is a critical need for high-quality clinical outcome assessments to capture the important aspects of communication ability of individuals with Angelman syndrome (AS). To center the perspective of caregivers, our team developed the novel Observer-Reported Communication Ability (ORCA) measure using best practice guidelines, with the goal of developing a measure that could be administered to caregivers directly without the need for a certified administrator for use in clinical trials. To refine the draft measure, we conducted two rounds of cognitive interviews with 24 caregivers and a quantitative study including 249 caregivers. The results from both studies support the overall content validity, construct validity, and the reliability of the ORCA measure for individuals with AS ≥ 2 years old for use in research contexts. Future work should explore the responsiveness of ORCA measures to changes over time in a diverse sample.

Angelman syndrome (AS) is a rare, neurodevelopmental disorder that significantly impacts the quality of life of the individual and their family through deficits in a number of areas, including communication (Wheeler, Sacco, & Cabo, 2017). For individuals with AS, communication is characterized by minimal to no speech production, limited expressive vocabulary, and greater receptive versus expressive language ability (N. Jolleff & Ryan, 1993; Keute et al., 2020; Pearson, Wilde, Heald, Royston, & Oliver, 2019; Penner, Johnston, Faircloth, Irish, & Williams, 1993). Individuals with AS generally use a wide variety of communication modalities including non-symbolic (gestures, vocalizations, physical manipulation of others), symbolic (signs, speech), and low- and high-tech augmentative and alternative communication (AAC) devices (Pearson et al., 2019). As a primary domain of daily functioning, communication is extremely important, as it directly impacts quality of life, socialization, safety, caregiving, and clinical care. Thus, it is not surprising that caregivers of individuals with AS have identified communication as a key, family-centered indicator of treatment efficacy for clinical trials (Willgoss et al., 2020). It is also ethically imperative to support the development of children's ability to communicate wants, needs, and ideas as effectively and efficiently as possible (Sigafos et al., 2000). This includes requesting, refusing, understanding others, and connecting with people in their environment.

Existing communication measures have rigorous validity evidence in certain contexts, however, none were developed using input from caregivers and families of individuals with AS and most require a trained clinician to administer and interpret scores. Although useful for their standardization and normative data, assessments designed primarily for identifying developmental delays and guiding clinicians in developmental care plans leave many individuals

with AS scoring at the floor of these measures, masking meaningful changes in communication ability (see Keute et al, 2020 for an illustrative example). Additionally, individuals with AS function poorly in unknown environments with strangers due to dyspraxia/apraxia (Greico, Bahr, Schoenberg, et al, 2016) and a focus on verbal speech does not cover the different modalities utilized by these individuals; thus even a well-administered ClinRO or PerfO measure is unlikely to completely capture communication ability in individuals with AS (Grieco et al, 2018; Penner et al., 1993; Cousins & Smyth, 2005). Thus, the field could benefit from the addition of a high quality caregiver-reported measure of communication ability for use alongside other clinician-reported outcome (ClinRO) and performance based outcome (PerfO) measures. Ideally, the content of this measure would focus on meaningful and relevant aspects of an individual's communication ability from the perspective of the family, be sensitive enough to detect change over time, and allow for independent administration via caregiver self-report.

To meet this need, our team developed the Observer-Reported Communication Ability (ORCA) measure for the AS population using best practice guidelines provided by the FDA, ISPOR, ISOQOL, PROMIS, and other organizations (DeWalt, Rothrock, Yount, & Stone, 2007; US Food & Drug Administration, 2009; Reeve et al., 2013; Rothman et al., 2009). The first step in this process was to conduct concept elicitation interviews with relevant stakeholders to identify the important components of communication and identify the language caregivers' use when describing their child's communication. We integrated this qualitative data from caregivers of individuals with AS and clinicians, including speech-language pathologists (SLPs), who have experience working with these families, to draft the ORCA measure. The goal of this manuscript is to evaluate the content validity and psychometric properties of the draft ORCA measure in a

sample of caregivers of individuals with AS through two phases: cognitive interviewing and a cross-sectional validation study. We used cognitive interviewing (commonly called ‘debriefing’) to ensure appropriateness of the items and coverage of the important aspects of the construct (in this case ‘communication ability’; Willis, 2005) and to ensure understanding of all components on the measure (e.g., instructions, items, response options, and recall period; Patrick et al., 2011). Our cross-sectional study was designed to quantitatively evaluate the psychometric properties of the measure, including reliability and construct validity (via comparing the ORCA scores to scores with other caregiver-reported measures). Our core study team included measurement methodologists and SLPs, and we were guided by parents and relatives of children with AS via active engagement in regular research team meetings. These individuals are also listed authors on this publication (A.B., J.P., & P.E). The overall goal of our work is to design a caregiver-reported measure that can be used in longitudinal research studies to examine an individual’s change over time in communication ability.

Methods

Experimental Design

Following established guidelines for developing clinical outcome assessments (Reeve et al., 2013; U. S. FDA, 2006), we conducted a two-phase exploration of the psychometric properties of the novel ORCA measure. Phase 1 involved cognitive interviewing with caregivers of children with AS to support content validity (Mokkink, Terwee, Patrick, et al., 2010), and Phase 2 involved a quantitative psychometric evaluation to provide evidence of construct validity and

reliability. Both phases were approved by the relevant Institutional Review Board, and the study conformed to standards in the US Federal Policy for the Protection of Human Subjects.

The Observer-Reported Communication Ability (ORCA) measure

The ORCA measure is a novel measure designed to assess a child's typical communication ability over the past 30 days from the perspective of their primary caregiver. The ORCA measure includes concepts across expressive, receptive, and pragmatic forms of communication (Table 1). The measure was designed to be completed by a caregiver independently, without the need for a trained administrator present (e.g., SLP or other clinician). The initial content was based off of qualitative interviews with caregivers and clinicians (reference companion article) and published work in early communication development (Oller, Eilers, Neil, et al. 1999; Kaiser & Roberts, 2011; Beuker, Rommelse, Donders, et al., 2013), supporting the meaningfulness and relevance to families.

Prior to cognitive interviewing, the initial version of the ORCA measure consisted of 80 questions/items and the recall period was "currently." Items were organized into communication concepts (e.g., refusal, requesting, responding to their name) and asked about observable communication behaviors that were identified by caregivers in the concept elicitation study. This 'concept by behavior' structure is similar to other existing measures of communication ability (e.g. The Communication Matrix, designed for clinical care contexts, Quinn & Rowland, 2017; Rowland & Fried-Oken, 2010). Draft response options were developed from the language caregivers used in concept elicitation interviews when describing their child's communication skills. Please see Table 1 for an example item set for the concept of 'refusal'.

A set of communication-specific descriptive items were also collected in tandem to capture important and detailed information about the individual's unique ways of communicating, including their vocabulary within each modality (e.g., gestures/signs, sounds, words/word approximations, etc.) and complexity of messaging (e.g., combining two symbols together to communicate one message). Although supported through the initial qualitative work, all of the aspects on the initial draft version of the ORCA measure (e.g., instructions, item wording, response options, and recall period) were intended to be evaluated within this study and refined as needed.

Cognitive Interviewing

Participants and Recruitment

In-depth cognitive interviews were conducted with adult caregivers of individuals with AS who were able to read, speak, and understand English. Their child had to be at least two years of age and caregivers had to report molecular confirmation of their child's diagnosis (e.g., DNA methylation test, FISH, CGH, or sequencing). Caregivers also had to live with the individual with AS and reside in the United States. Recruitment of caregivers was stratified by child age, using categories of 2-7 years, 8-12, 13-17, and ≥ 18 years, to ensure representation of communication ability across the lifespan. We also sought heterogeneity among the caregivers in terms of their child's AS genotype: deletion positive, mutation/*UBE3A*, imprinting center defect (ICD), or uniparental disomy (UPD). Purposeful sampling was used across genotype to closely represent the national prevalence (as per Clayton-Smith & Laan, 2003). A recruitment flyer was posted on the Facebook page of the Foundation for Angelman Syndrome Therapeutics (FAST), a

patient advocacy organization, and was subsequently circulated within other similar forums (parent support groups, etc.).

Our team conducted two rounds of cognitive interviewing with 24 caregivers, with 12 in each round. Interviews were conducted between June and October, 2019. Participants received a \$50 gift card for participating. A majority of the caregivers were mothers and white (Table 2), with an almost even split between male and female individuals with AS. The majority of individuals with AS had deletion positive genotype (58%), matching national prevalence (Clayton-Smith & Laan, 2003).

Procedures

Two rounds of cognitive interviewing were performed using retrospective probing techniques (as described in Willis, 2004; Willis, 2005). The target sample size of 12 caregivers per round (3 per age group) meets the recommended sample size for cognitive interviewing (Willis, 2004).

Individual interviews were designed to last 60-90 minutes, and were conducted over the phone. Prior to the interview, study personnel mailed a paper copy of the ORCA measure to the participant with instructions to wait to complete it until the time of the interview. On the call, after obtaining verbal consent, trained interviewers began by providing a brief overview of interview procedures and then asked the participant to complete the ORCA measure. Participants were instructed to circle any questions or words that were hard to understand and complete the measure independently. After completing the ORCA measure, the interviewer then used structured probes to ask about understandability, readability, content, over- and under-representation of the items, and performance of the overall measure. Interviewers also evaluated

the appropriateness of the recall period, response options, and instructions. All interviews were audio recorded and transcribed. Interviewers completed a debriefing form after each interview to document and summarize main findings.

Between the first and second round of interviews, the updated version of the ORCA measure was sent to the FACITrans Company to review the translatability of the measure into different languages (e.g., Spanish, Chinese). The company provided the study team with a line-by-line review of the ORCA measure, which included a difficulty rating (0=no issues, 1=minor difficulty to translate, 2=moderate difficulty, 3=a lot of difficulty, and 4=impossible to translate). For ratings of 1-3, alternative wordings were provided by the company to improve the likelihood of success in future language translations. This step is recommended as a way to evaluate if a measure can be meaningfully translated in the future (Acquadro, Patrick, Eremenco, et al., 2017), but it is not a full translation process, nor is it a formal evaluation of cross-cultural validity.

Analysis

After each interview, interviewers made note of any problems or issues with the ORCA measure's instructions, items, and response options, and documented their impressions in structured debriefing forms. After the first round of interviews was completed ($n = 12$), team members reviewed debriefing forms and subsequently made decisions on what parts of the ORCA measure required revision. Revisions were discussed among team members, who included measurement experts, SLPs, and caregivers of children with AS. Items were defined as being "significantly revised" if their revision involved (1) adding or removing a word(s) that changed the meaning of a phrase, (2) word substitutions that in the judgment of the investigators

were more than a semantic simplification, or (3) significant changes to the response options (e.g., changing from a severity to a frequency scale). Significant changes made after Round 1 (including those indicated by the translatability review) were evaluated in Round 2, and an item-tracking matrix was used to capture all revisions and justifications for changes.

Readability. We estimated the reading level of the modified text using online software (Readable) to obtain the Flesch-Kincaid Grade Level (Kincaid, Fishburne, Rogers, & Chissom, 1975), which provides a numeric score representing ‘readability’ of text that can be interpreted as the United States grade level of education. Consensus is generally that medical materials for adult patients, including outcome measures, should be targeted to a 6th grade reading level or below (Eltorai, Ghanian, Adams, Born, & Daniels, 2014; Douglas & Kelly-Campbell, 2018; Wilson, 2009).

Psychometric Testing

Participants

Eligibility criteria for caregivers enrolled in the psychometric study was the same as for cognitive interviewing with the additional criteria that the individual with AS in their care had to be between the ages of 2-40 years old. The age range was broad to encompass communication abilities from a large group of individuals with AS. Although not specifically recruited, participants who participated in the previous qualitative work were eligible to participate in this study.

Validation of the Observer-Reported Communication Ability (ORCA) Measure for Individuals with Angelman Syndrome

The electronic survey was initiated by 295 caregivers of individuals with AS, with 249 having adequate completeness on their responses on the ORCA measure and thus were included in analyses. The majority of caregivers were women (88.2%) who were on average 41.6 years old (Table 2). Individuals with AS ranged in age from 2 to 39 years, were an average age of 10.5 years, and there was representation from all genotypes (Table 2). Caregiver education varied, with one reporting less than high school diploma (0.4%), 24 reporting high school degree or equivalent (9.7%), 58 reporting some college/university (23.5%), 105 reporting college/university degree (42.5%), and 59 reporting a postgraduate degree (23.9%). Family income also varied; 50 reported annual household income between 0 and \$60,000 (20%), 27 between \$60,000 and \$80,000 (10.8%), 39 between \$80,001 and \$100,000 (15.7%), 82 between \$100,001 and \$250,000 (32.9%), and 26 over \$250,000 (10.0%; 23 caregivers chose not to answer this question; 9.2%).

Procedures

Recruitment was remote and included a web link posted on the patient advocacy organization's Facebook page. Interested participants could click on the link and complete the survey via the Research Electronic Data Capture (REDCap) system. The REDCap survey was active from December 6, 2019 through January 10, 2020. Participants were not compensated for participating in this portion of the study.

After providing informed consent, the survey included demographic information from the caregiver about the individual with AS including seizure activity, recent illness, and hospitalizations. Participants were asked to enter their email address in order to receive an email

reminder 5 to 12 days later to complete the ORCA measure a second time (to be used to estimate test re-test reliability).

Measures

Additional measures were also collected to evaluate construct validity. The Communication and Symbolic Behaviors Scale DP™ Infant-Toddler Checklist (CSBS; Wetherby & Prizant, 2003; Wetherby, Allen, Cleary, Kublin, & Goldstein, 2002) was chosen as a corresponding measure of communication ability. The CSBS was developed as a tool to screen children for delays in early social communication (Wetherby & Prizant, 2003) and has published support for validity when used with young children (6-24 months) for this purpose. It was chosen for this study due to its short length (24 items) and because it was specifically developed for caregivers to complete independently without a trained administrator present. It was not designed specifically for older children, however, it does include non-verbal communication behaviors like gestures, facial expressions, and positive affect, which are commonly used by individuals with AS. The CSBS has been previously utilized in AS samples, although published validity evidence has been quantitative and not focused on content validity (Hamrick & Tonnsen, 2019; Hamrick, Haney, Kelleher, & Lane, 2020). Higher CSBS scores represent better communication ability.

To further evaluate construct validity, the study team also included measures of mobility (the PROMIS® Parent Proxy Physical Function–Mobility 8-item short form; PROMIS-PF; Irwin et al., 2012) and sleep disturbance (the PROMIS® Parent Proxy Sleep Disturbance 8-item short form; PROMIS-Sleep; Forrest et al., 2018). We expected a moderate but positive association between scores on the ORCA and mobility measures. We did not have strong *a priori* hypotheses

about the association of sleep and communication, although we expected correlations to be weaker in magnitude when compared to mobility, and thus, it was considered an ‘exploratory’ indicator. The recall period for both these measures was “the past 7 days”.

Statistical Analyses

We explored structural validity, internal consistency and test-retest reliability, presence of floor/ceiling effects, known-groups validity, construct validity, and an exploration of potential minimal clinically important differences. Details for each analysis are described below.

Structural validity. Based on our conceptual framework, we anticipated an overall communication ability construct that included concepts within expressive, receptive, and pragmatic forms of communication. Thus, we used confirmatory factor analysis (CFA) for categorical response data using weighted least squares means and variance adjusted (WLSMV) estimation (Beauducel & Herzberg, 2006) implemented in MPLUS software [version 8.2]. Fit statistics were evaluated and included the comparative fit index (CFI; $>.95$ for very good fit), Tucker-Lewis Index (TLI; $>.95$ for very good fit), root mean square error of approximation (RMSEA, $<.06$ for very good fit; McDonald, 1999).

Score Distribution and Reliability. Following confirmation of the factor structure, the ORCA measure was scored based on the caregiver’s assessment of an individual’s ability to “master” communication behaviors leveled within each of the communication concepts reflective of expressive (e.g., seek attention), receptive (e.g., making choices), and pragmatic (e.g., greeting) forms of communication. Mastery was defined as an individual performing the behavior

frequently and consistently over the past 30 days, aligned with the wording of the response options and specified recall period. Graded response item response theory (IRT) modeling was used to calibrate the scale, and the model adjusted for the discrimination and difficulty levels of the concepts (Samejima, 1969). Thus, IRT modeling was applied to the 23 communication concepts within the ORCA measure; individual concepts potentially included multiple questions on the ORCA measure. For example, the concept “Respond to Familiar Directions” includes the caregiver’s answers to four ORCA questions related to their child’s mastery of “stopping” something (question 17a), following “one-step directions” (question 17b), “two-step directions” (question 17c), or “three-step directions” (question 17c), all which indicated differing levels of skill within the concept. IRT scoring used the *expected a posteriori* (EAP) estimate based on the response patterns observed in the data, with scores transformed to a T-score metric with mean 50 and standard deviation (SD) of 10.

We reported reliability in terms of: 1) the extent to which the ORCA measures’ IRT-based information function shows high information across the range of participant scores; 2) Cronbach’s coefficient alpha for internal consistency (threshold $> .70$ for group level assessment; Reeve et al., 2013; Terwee et al., 2007); and 3) test-retest reliability using a two-way mixed effects model on absolute agreement to calculate interclass correlation coefficients (thresholds of $.70$; Koo & Li, 2016). We also evaluated the percent of individuals who scored at the lowest (floor) and highest (ceiling) possible ORCA score.

Known-groups validity. We examined difference in ORCA score among AS genotypes (deletion positive, mutation, UPD, and ICD) using a one-way between-group ANOVA and conducted

pairwise comparisons on group least-square means. We expected individuals with deletion positive genotype to have lower communication ability scores on the ORCA measure compared to individuals with other genotypes (Didden, Korzilius, Duker, & Curfs, 2004; Nicola Jolleff, Emmerson, Ryan, & McConachie, 2006; Keute et al., 2020). We also compared ORCA scores between individuals with and without seizures using an independent samples t-test. We expected children that experienced a seizure in the past year to have lower communication ability scores on the ORCA measure than those who had not.

Age & ORCA scores. To explore the relationship between age and ORCA scores, we estimated the Pearson correlation coefficient. Our hypothesis was that there would be small to moderate positive correlations between communication ability and age, as significant variability was expected in ability across children at similar chronological ages. To further explore score ranges within age cohorts, we also report mean, standard deviations, and minimum/maximum for children 2-4 years old, 5 to 12 years old, and 13 to 35 years old.

Construct validity. We estimated Pearson correlations between the ORCA scores and scores on the CSBS, PROMIS-PF and PROMIS-Sleep. Correlations were classified as: small (0.10–0.29), moderate (0.30–0.49), strong (0.50–0.69), and very strong (>0.70; J. Cohen, 1992; Jacob Cohen, 2013). We expected the ORCA to have very strong correlations with the CSBS, moderate to strong correlations with PROMIS-PF, and small correlations with PROMIS-sleep.

Exploration into Minimal Important Differences. We also performed an exploratory analysis of possible minimal important differences (MID) estimates. Using a distribution-based approach,

we compared score differences between known groups to standard effect sizes (Cohen's effect size between .2 [small] and .5 [moderate]). Although this data is cross-sectional, previous work suggests that roughly a 2 to 5 point difference (i.e., .2 to .5 SDs) may represent a MID (Norman, Sloan, & Wyrwich, 2003). Although this provides an initial estimate of potential MIDs that could be detected, future work using longitudinal data and anchor based methods would be more appropriate to define what is minimally 'important' (de Vet, Terwee, Ostelo, et al., 2006).

Results

Cognitive Interviewing

Round 1

After the first round of cognitive interviewing, data suggested the addition of a more specified recall period and changes to the response options. Justification around these changes are discussed below. A detailed description of all modifications made across both rounds of interviews can be found in Appendix 1.

Recall period. The first version of the ORCA measure purposely did not have any specified recall period other than "currently" so that interviewers could probe on the period caregivers automatically used when considering their child's communication ability. The vast majority of caregivers reported thinking over a long period of time (a month or more) to reflect 'typical' communication. When asked about an appropriate recall period, caregivers reported that one week of communication would not be the best representation of ability as their child's communication could be influenced by changes in routine, illness, or seizure activity. After

carefully considering characteristics of communication in these individuals, accurate recall, and potential data collection within a future clinical trial (Norquist, Girman, Fehnel, DeMuro-Mercon, & Santanello, 2012), the study team decided to set the recall period at 30 days and evaluate this addition in Round 2. The study team hypothesized that 30 days would be long enough that someone's communication would not be completely overwhelmed by short term issues (e.g. illness), but short enough that the caregiver could accurately recall their child's behavior.

Response options. When evaluated in the first round of cognitive interviewing, caregivers reported interpreting "consistently" in varied ways, with most of them translating them to frequencies. For example, one participant stated, "I would say ['yes, but not consistently'] if he grabbed but only some of the time or occasionally....if it's something their child does every day or their go-to method, then I would say 'yes, consistently.'" Based on feedback from caregivers and through extensive discussions with the study team, the response options were changed to "No or only once," "Sometimes," and "Yes, almost all the time," and evaluated again in Round 2. These response options were designed to represent skill mastery ("Yes, almost all the time"), meaning individuals could and did perform the behavior frequently and consistently, and emerging skills ("sometimes"), meaning individuals may be starting to use these behaviors more consistently, but not to the same extent as mastered skills.

Translatability review. The final report on the ORCA measure showed no major expected issues with future language translations; no wording or phrasing were deemed impossible or extremely difficult to translate. Words deemed "moderately difficult" to translate were changed based on

the FACITtrans' recommendations prior to Round 2 of cognitive interviewing so they could be evaluated. Notable changes included changing all pronouns to the third person singular (e.g., from "they" to "he/she") in items referring to the child.

Round 2

During the second round of cognitive interviewing, interviewers confirmed that questions on the key concepts (e.g., seeking attention, refusing an object) captured the types of communication behaviors caregivers had observed in their children, and that the recall period of 30 days was appropriate. Interviewers also asked caregivers to share examples of these behaviors, and feedback from caregivers was consistent with the concept being measured by the ORCA measure.

Round 2 feedback indicated that the new frequency response options reflected the caregivers' personal conceptualizations of "mastery" in communication skills. In agreement with our formative work (please see companion article), caregivers reported confidence that their child had mastered a skill when they used it frequently, consistently, and in/across different settings (e.g. home and school). Based on their feedback, the response options of "No or only once," "Sometimes," and "Yes, almost all the time" reflected these observations and generally represented the lack of a skill, the emergence of a skill, and mastery of a skill (respectively). Although there was still some heterogeneity in how participants reported interpreting the response options (particularly around the interpretation of "sometimes"), the study team felt that overall, the new response options performed better than previous options.

Reading level. The estimated Flesch-Kincaid grade level was 5, corresponding to a 5th grade reading level. The most advanced words that were identified were ‘approximations’ (referring to ‘word approximations’) and ‘conversation’.

Psychometric Testing

Structural Validity. Based on item fit statistics, two concepts and their related items were identified for possible removal (“telling about the past,” “telling about feeling sick”). The concept “telling about feeling sick” had low associations with other communication ability concepts and was removed by the research team without impacting content validity of the ORCA measure (see discussion section). The concept “telling about the past” represented a very high ability item that was infrequently reported by caregivers, thus model estimates for this item were not stable. A one-factor CFA model including the remaining concepts found evidence for very good model fit; CFI = 0.96; TLI = 0.95; RMSEA = 0.06, 90% CI: 0.05-0.07, supporting the reporting of one total score to represent overall ‘communication ability’ across the included concepts.

Score Distribution and Reliability. The distribution of ORCA total scores was roughly bell-shaped (Figure 1). Only 3 participants (1.2%) had the lowest possible ORCA score (T-score = 26.8) and no one was at the ceiling (highest possible = 83.24; highest in sample = 76.4). The IRT-based information function showed high information across the range of participant scores with most of the information aligning with the mean of communication ability (Figure 2). Internal consistency (Cronbach’s coefficient alpha) was 0.90. Among the 170 caregivers who

Validation of the Observer-Reported Communication Ability (ORCA) Measure for Individuals with
Angelman Syndrome

completed the ORCA measure at both time points, the test re-test reliability was excellent (ICC = 0.91; 95% CI: 0.88-0.93).

Known-groups validity. Individuals with Deletion Positive genotype had communication ability scores that were significantly lower than the other genotypes (Table 3). Average scores for the other three genotypes were not statistically different from each other. Additionally, children that experienced a seizure in the past year had significantly lower ORCA scores ($mean = 48.5, sd = 9.7$) than children who had not experienced a seizure in the past year ($mean = 51.9, sd = 9.2; t = 2.86, df = 246, p < .01$).

Age & ORCA scores. The correlation between age and the ORCA score was small, but statistically significant ($r = 0.15; p = 0.021$). Average scores within each age group varied, with significant overlap in score range across age categories (Table 3).

Construct Validity. Total scores on the ORCA measure were strongly correlated with the CSBS total score ($r = .83, p < .001$). A moderate-strong correlation was found between ORCA score and PROMIS-PF measure ($r = .53, p < .001$). No association was found between the ORCA measure and the PROMIS-Sleep Disturbance measure ($r = -.09, p = .15$). There also was no association between PROMIS-Sleep Disturbance scores with the CSBS communication measure ($r = -.07, p = .31$).

Minimal important differences. Mean differences in scores between individuals with deletion positive and other genotypes in our sample ranged from 10.4 to 14.0, which exceeded the threshold of 5 (.5 SD).

DISCUSSION

The ORCA measure is a communication ability measure designed for use in clinical trials and centering the perspectives of the caregivers of individuals with AS. The development work was a partnership between advocacy, caregivers, and academia, with the goal of creating a tool that would capture this clinically meaningful domain with this unique patient community in mind. The results from cognitive interviewing and the psychometric evaluation support the overall content validity, construct validity, and the reliability of the measure for this use. While some expected refinement of concepts occurred within the two-phased evaluations, overall, the content remained stable.

As the ORCA measure was specifically designed for use in clinical trials for individuals with AS, we paid particular attention to two aspects of the measure during cognitive interviewing: the recall period and the response options. The recall period was intentionally left ambiguous in the first round of interviews so that we could explore the timeframes that caregivers naturally considered when answering the questions, without undue influence by the study team. Based on caregiver feedback, the final 30-day recall period was long enough to capture a range of typical communication behaviors without reflecting fluctuations in communication based on illness or schedule changes. Additionally, the final recall period was long enough to capture stable and

meaningful changes in mastered communication skills while allowing caregivers confidence in their recall of each observable communication behavior.

The response options were modified after the first round of cognitive interviewing to better reflect how caregivers discussed emerging vs mastered communication skills. There was heterogeneity in how caregivers were interpreting the initial response options, and some reported confusion on how to map them to concepts. Since most of the caregivers were interpreting the original response options (“consistently” versus “not consistently”) in terms of frequency, this language was used directly.

The psychometric results also provided strong quantitative support for the ORCA measure and the reporting of one total score representing overall communication ability. The score distribution was not skewed and there were minimal floor effects (i.e., only 3 individuals were at the lowest score) and no ceiling effects in this sample. This is an improvement over existing communication measures that are used with individuals with AS, and supports the idea that this measure is uniquely tailored to the communication ability levels seen within this population. High quality outcome measures are critical for clinical trials, as we are less likely to identify positive shifts in the distribution when it is skewed or there is minimal variability within a sample. There were also minimal concerns based on internal stability, reliability, or construct validity.

In terms of content, the psychometric results identified two communication concepts that needed further review by the study team. First, the concept “telling about feeling sick” had poor item

performance. This item was originally included in the ORCA measure based on results from our formative work; caregivers and SLPs identified this as a potentially meaningful skill for children in this context (please see companion article). However, the study team also recognized that these concepts were not always relevant to every family, depending on their priorities for their child and skills that they were targeting/reinforcing. Thus, when it showed poor item performance, the study team felt comfortable that removing these items would not impact the overall content validity of the ORCA measure. Children who can inform caregivers of pain and illness have higher level expressive and receptive skills, and these should be captured through other relevant and better-performing items (e.g. directing attention to a body part, answering yes/no questions like ‘does your head hurt’). The second concept that the item level statistics identified was “telling about the past.” This represented a very high ability item that caregivers infrequently reported seeing in their children, thus model estimates for this item were not stable. However, we believed that it was important for content validity to continue to capture this advanced level skill, especially in the context of a clinical trial when improvement in communication is the goal. Thus, at this stage of measure development, we recommend that this item remain as part of the ORCA measure, but will not contribute to the total score without further evaluation.

Although our samples were well representative of age cohorts and AS genotypes, the majority of caregivers in our sample were white and had at least some college or graduate education. This is consistent with the current underrepresentation of minority populations in genetic studies and reflects systemic structural barriers that can limit or delay diagnosis and treatment for rare diseases, especially for communities of color (D'Angelo et al., 2020; Fraiman & Wojcik, 2020).

Future work should continually ensure the ORCA measure's content and psychometric properties are equally supported when used with all families, including those who are most strongly impacted by institutionalized and structural racism. We also limited our recruitment to English-speaking families in the continental United States and required parents to confirm that their child's diagnosis was based on genetic testing. The latter is not available to all families, so it could represent a financial barrier to participation. We also limited the sample to children who were 2 years old and older, and thus, do not recommend the use of the ORCA measure in children below this threshold.

The current work, although supportive of the modified ORCA measure, does not fully address all aspects of reliability and validity. Future studies are planned to explore the relationship of the ORCA measure with other performance-based or clinician-reported communication measures, like the Vineland Adaptive Behavior Scales (Sparrow, Cicchetti, & Balla, 2005), as well as longitudinal validity in existing developmental cohorts and after known interventions. We also recommend a deeper exploration of the relationship between scores and age, as normative values would be useful for interpreting clinical trial results, and between scores and cognitive ability (which was not assessed in this study). We also note that the ORCA measure was developed specifically for individuals with AS, but the measure may have applications to other neurodevelopmental disorders that have similar communication profiles.

Conclusion

The inclusion of high quality and meaningful clinical outcome measures is crucial to ensure upcoming clinical trials capture the outcomes that matter to individuals with AS and their

families. The novel ORCA measure was designed to capture expressive, receptive, and pragmatic communication ability for individuals with AS ≥ 2 years of age from the perspective of their caregiver when completed independently, without the need for a trained administrator. The results of our study provides initial evidence supporting the validity and reliability of this measure in this context of use, although evaluations of the responsiveness over time due to changes in communication with diverse samples are still needed. We believe that the ORCA measure can provide meaningful and useful information about the communication ability of an individual with AS when included in future clinical trials. In particular, the addition of a disease-specific caregiver-centered measure alongside other performance-based and clinician-reported measures can provide a comprehensive assessment of the child's skills and changes over time in response to treatment through the eyes of experts and their family.

References

- Acquadro, C., Patrick, D. L., Eremenco, S., Martin, M. L., Kuliś, D., Correia, H., Conway, K., & International Society for Quality of Life Research (ISOQOL) Translation and Cultural Adaptation Special Interest Group (TCA-SIG) (2017). Emerging good practices for Translatability Assessment (TA) of Patient-Reported Outcome (PRO) measures. *Journal of patient-reported outcomes*, 2(1), 8. <https://doi.org/10.1186/s41687-018-0035-8>
- Beauducel, A., & Herzberg, P. Y. (2006). On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural equation modeling: a multidisciplinary journal*, 13(2), 186-203.
doi:10.1207/s15328007sem1302_2
- Beuker, K. T., Rommelse, N. N., Donders, R., & Buitelaar, J. K. (2013). Development of early communication skills in the first two years of life. *Infant Behavior and Development*, 36(1), 71-83.
- Clayton-Smith, J., & Laan, L. (2003). Angelman syndrome: a review of the clinical and genetic aspects. *J Med Genet*, 40(2), 87-95. doi:10.1136/jmg.40.2.87
- Cohen, J. (1992). A power primer. *Psychol Bull*, 112(1), 155-159. doi:10.1037//0033-2909.112.1.155
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*: Academic press.
- Cousins M, & Smyth, M. M. (2005). *Progression and development in Developmental Coordination Disorder. In: Children with Developmental Coordination Disorder*. London, UK.: Whurr Publishers, pp. 119-134.
- D'Angelo, C. S., Hermes, A., McMaster, C. R., Prichep, E., Richer, É., van der Westhuizen, F. H., . . . Baynam, G. (2020). Barriers and Considerations for Diagnosing Rare Diseases in

Validation of the Observer-Reported Communication Ability (ORCA) Measure for Individuals with
Angelman Syndrome

Indigenous Populations. *Frontiers in pediatrics*, 8, 579924-579924.

doi:10.3389/fped.2020.579924

de Vet, H. C., Terwee, C. B., Ostelo, R. W., Beckerman, H., Knol, D. L., & Bouter, L. M.

(2006). Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and quality of life outcomes*, 4(1), 1-5.

DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. (2007). Evaluation of item candidates: the PROMIS qualitative item review. *Med Care*, 45(5 Suppl 1), S12-21.

doi:10.1097/01.mlr.0000254567.79743.e2

Didden, R., Korzilius, H., Duker, P., & Curfs, L. (2004). Communicative functioning in individuals with Angelman syndrome: a comparative study. *Disabil Rehabil*, 26(21-22), 1263-1267. doi:10.1080/09638280412331280271

Eltorai AE, Ghanian S, Adams CA Jr, Born CT, Daniels AH. Readability of patient education materials on the American Association for Surgery of Trauma website. *Arch Trauma Res*. 2014;3(2):e18161. Published 2014 Apr 30. doi:10.5812/atr.18161

Forrest, C. B., Meltzer, L. J., Marcus, C. L., de la Motte, A., Kratchman, A., Buysse, D. J., . . .

Bevans, K. B. (2018). Development and validation of the PROMIS Pediatric Sleep Disturbance and Sleep-Related Impairment item banks. *Sleep*, 41(6).

doi:10.1093/sleep/zsy054

Fraiman, Y. S., & Wojcik, M. H. (2020). The influence of social determinants of health on the genetic diagnostic odyssey: who remains undiagnosed, why, and to what effect?

Pediatric Research. doi:10.1038/s41390-020-01151-5

Validation of the Observer-Reported Communication Ability (ORCA) Measure for Individuals with
Angelman Syndrome

Grieco, J.C., Bahr, R.H., Schoenberg, M.R., Conover, L., Mackie, L.N., Weeber, E.J. (2018).

Quantitative Measurement of Communication Ability in Children with Angelman
Syndrome. *J Appl Res Intellect Disabil*; 31: e49-e58.

*Guidance for industry: patient-reported outcome measures: use in medical product development
to support labeling claims.* (2009). Retrieved from

<https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>.

Hamrick LR, Tonnsen BL. Validating and Applying the CSBS-ITC in Neurogenetic Syndromes.

Am J Intellect Dev Disabil. 2019;124(3):263-285. doi:10.1352/1944-7558-124.3.263

Hamrick, L.R., Haney, A.M., Kelleher, B.L., & Lane, S.P. (2020). Using generalizability theory

to evaluate the comparative reliability of developmental measures in neurogenetic

syndrome and low-risk populations. *Journal of Neurodevelopmental Disorders*; 12:16.

<https://doi.org/10.1186/s11689-020-09318-1>

Irwin, D. E., Gross, H. E., Stucky, B. D., Thissen, D., DeWitt, E. M., Lai, J. S., . . . DeWalt, D.

A. (2012). Development of six PROMIS pediatrics proxy-report item banks. *Health Qual*

Life Outcomes, 10, 22. doi:10.1186/1477-7525-10-22

Jolleff, N., Emmerson, F., Ryan, M., & McConachie, H. (2006). Communication skills in

Angelman Syndrome: Matching phenotype to genotype. *Advances in Speech Language*

Pathology, 8(1), 28-33. doi:10.1080/14417040500459684

Jolleff, N., & Ryan, M. M. (1993). Communication development in Angelman's syndrome. *Arch*

Dis Child, 69(1), 148-150. doi:10.1136/adc.69.1.148

Kaiser A. P., Roberts M. Y. (2011). Advances in Early Communication and Language

Intervention. *Journal of Early Intervention*, 33 (4), 298-309.

doi:10.1177/1053815111429968

Validation of the Observer-Reported Communication Ability (ORCA) Measure for Individuals with
Angelman Syndrome

Keute, M., Miller, M. T., Krishnan, M. L., Sadhwani, A., Chamberlain, S., Thibert, R. L., . . .

Hipp, J. F. (2020). Angelman syndrome genotypes manifest varying degrees of clinical severity and developmental impairment. *Mol Psychiatry*. doi:10.1038/s41380-020-0858-6

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*, 15(2), 155-163.

doi:10.1016/j.jcm.2016.02.012

McDonald, R. (1999). Test theory: a unified treatment. Lawrence Earlbaum Associates. *Inc.*, Mahwah, NJ, 142-145.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., de Vet, H.C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*; 63(7):737-45. doi:

10.1016/j.jclinepi.2010.02.006.

Norman, G. R., Sloan, J. A., & Wywich, K. W. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*, 41(5), 582-592. doi:10.1097/01.MLR.0000062554.74615.4C

Norquist, J. M., Girman, C., Fehnel, S., DeMuro-Mercon, C., & Santanello, N. (2012). Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration.

Quality of Life Research, 21(6), 1013-1020. doi:10.1007/s11136-011-0003-8

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of communication disorders*, 32(4), 223-245.

Validation of the Observer-Reported Communication Ability (ORCA) Measure for Individuals with
Angelman Syndrome

- Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. *Value in health: The journal of the International Society for Pharmacoeconomics and Outcomes Research*, 14(8), 978–988.
<https://doi.org/10.1016/j.jval.2011.06.013>
- Pearson, E., Wilde, L., Heald, M., Royston, R., & Oliver, C. (2019). Communication in Angelman syndrome: a scoping review. *Dev Med Child Neurol*. doi:10.1111/dmcn.14257
- Penner, K. A., Johnston, J., Faircloth, B. H., Irish, P., & Williams, C. A. (1993). Communication, cognition, and social interaction in the Angelman syndrome. *Am J Med Genet*, 46(1), 34-39. doi:10.1002/ajmg.1320460108
- Quinn, E. D., & Rowland, C. (2017). Exploring Expressive Communication Skills in a Cross-Sectional Sample of Children and Young Adults With Angelman Syndrome. *Am J Speech Lang Pathol*, 26(2), 369-382. doi:10.1044/2016_ajslp-15-0075
- Readable [Computer software]. (2022). Retrieved from <https://readable.com/>
- Reeve, B. B., Wyrwich, K. W., Wu, A. W., Velikova, G., Terwee, C. B., Snyder, C. F., . . . Butt, Z. (2013). ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*, 22(8), 1889-1905. doi:10.1007/s11136-012-0344-y
- Rothman, M., Burke, L., Erickson, P., Leidy, N. K., Patrick, D. L., & Petrie, C. D. (2009). Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use

Validation of the Observer-Reported Communication Ability (ORCA) Measure for Individuals with
Angelman Syndrome

of Existing Instruments and Their Modification PRO Task Force Report. *Value Health*,
12(8), 1075-1083. doi:10.1111/j.1524-4733.2009.00603.x

Rowland, C., & Fried-Oken, M. (2010). Communication Matrix: A clinical and research
assessment tool targeting children with severe communication disorders. *Journal of
Pediatric Rehabilitation Medicine*, 3, 319-329. doi:10.3233/PRM-2010-0144

Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded
Scores.(Psychometrika Monograph, No. 17). Psychometric Society, Richmond.
Richmond, VA, USA.

Sigafoos, J., Woodyatt, G., Keen, D., Tait, K., Tucker, M., Roberts-Pennell, D., & Pittendreigh,
N. (2000). Identifying Potential Communicative Acts in Children with Developmental
and Physical Disabilities. *21(2)*, 77-86. doi:10.1177/152574010002100202

Sparrow, S. S., Cicchetti, D., & Balla, D. A. (2005). Vineland Adaptive Behavior Scales, Second
Edition (Vineland-II). *APA PsycTests*. doi:https://psycnet.apa.org/doi/10.1037/t15164-
000

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de
Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health
status questionnaires. *J Clin Epidemiol*, 60(1), 34-42. doi:10.1016/j.jclinepi.2006.03.012

U. S. Food & Drug Administration, Center for Drug Evaluation and Research, Center for
Biologics Evaluation and Research, and Center for Devices and Radiological Health.
(2006). Guidance for industry: patient-reported outcome measures: use in medical
product development to support labeling claims: draft guidance. *Health Qual Life
Outcomes*, 4, 79. doi:10.1186/1477-7525-4-79

US FDA Patient-Focused Drug Development Guidance Public Workshop on Guidance 4:

Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making. . Retrieved from <https://www.fda.gov/media/132505/download>

Wetherby, A. M., Allen, L., Cleary, J., Kublin, K., & Goldstein, H. (2002). Validity and reliability of the communication and symbolic behavior scales developmental profile with very young children. *J Speech Lang Hear Res, 45*(6), 1202-1218. doi:10.1044/1092-4388(2002/097)

Wetherby AM, & Prizant G (2003). *CSBS DP Manual. First Normed Edition*. Baltimore, MD: Brookes Publishing.

Wheeler, A. C., Sacco, P., & Cabo, R. (2017). Unmet clinical needs and burden in Angelman syndrome: a review of the literature. *Orphanet J Rare Dis, 12*(1), 164. doi:10.1186/s13023-017-0716-z

Willgoss, T., Cassater, D., Connor, S., Krishnan, M. L., Miller, M. T., Dias-Barbosa, C., . . .

Bichell, T. J. (2020). Measuring What Matters to Individuals with Angelman Syndrome and Their Families: Development of a Patient-Centered Disease Concept Model. *Child Psychiatry & Human Development*. doi:10.1007/s10578-020-01051-z

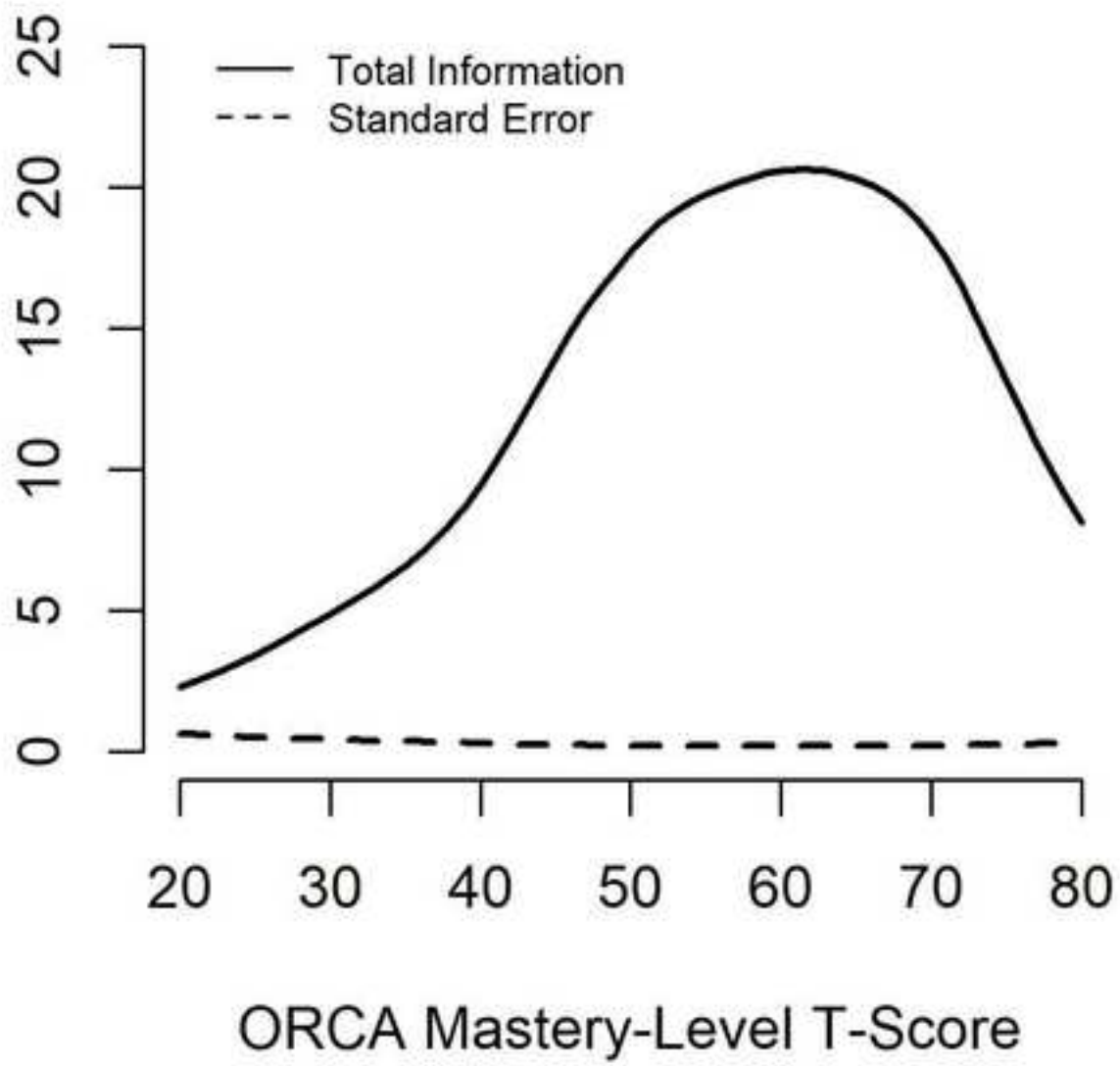
Willis, G.B. (2004). Cognitive interviewing revisited: A useful technique, in theory? *Methods for Testing and Evaluating Survey Questionnaires*. Ed. Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., & Singer. E. John Wiley & Sons Inc.

Willis, G. B. (2005). Cognitive interviewing in practice: think-aloud, verbal probing and other techniques. *Cognitive interviewing: a tool for improving questionnaire design*. London: Sage Publications, 42-63.

Wilson M. (2009). Readability and patient education materials used for low-income populations.

Validation of the Observer-Reported Communication Ability (ORCA) Measure for Individuals with
Angelman Syndrome

Clin Nurse Spec; 23: 33–40.



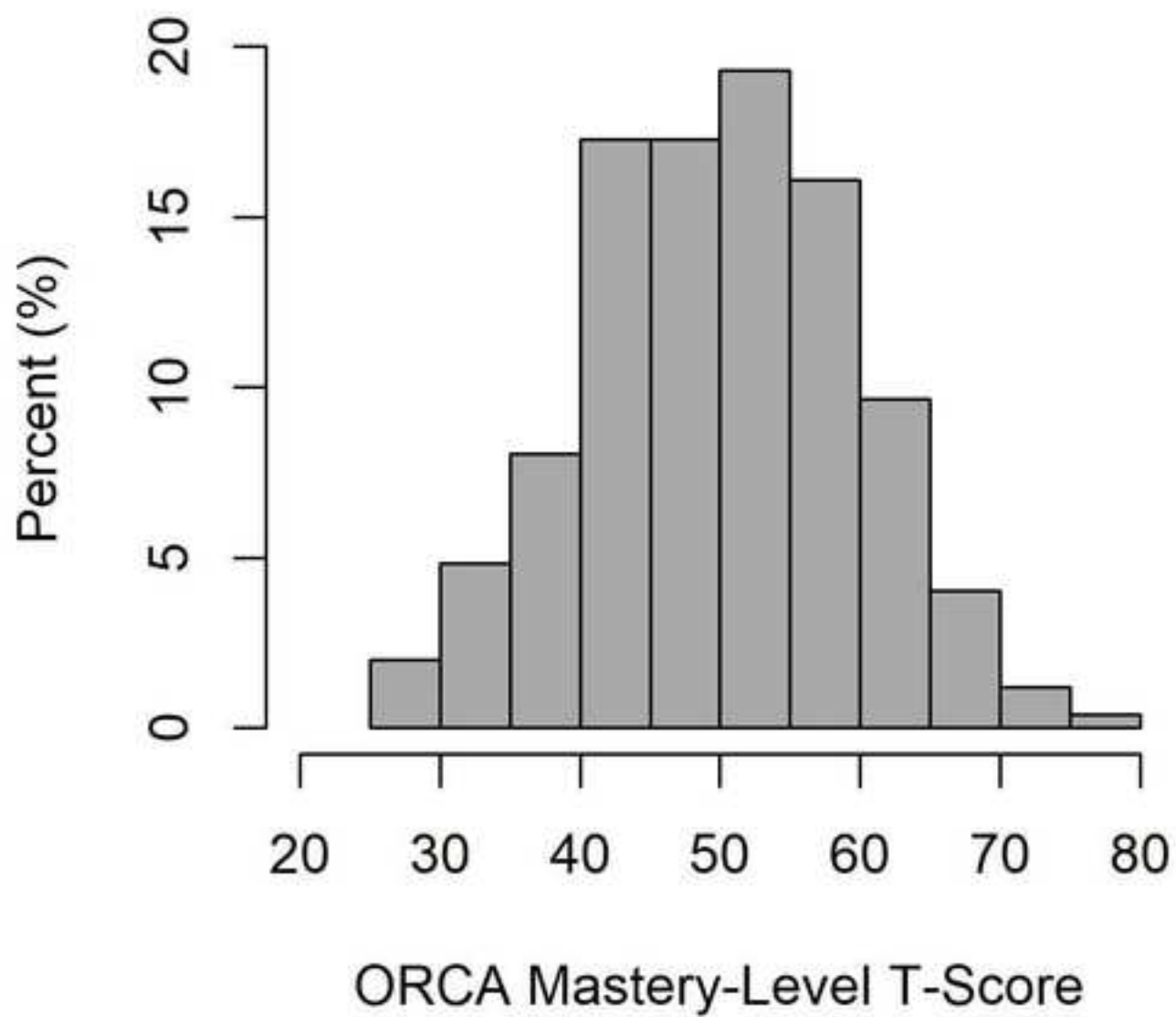


Table 1: The communication concepts included in the ORCA Measure and an example of a draft item set, showing behavioral items and the associated response options.

Expressive Communication	Receptive Communication	Pragmatic Communication		
Seek Attention	Respond to Name	Greeting		
Direct Attention	Understand Mood	Comfort Others		
Refuse Object	Understand Isolated Words	Play Games		
Request Object	Turns in Conversation	Use Names		
Request Object Out of View	Make Choices			
Request “More”	Respond to Familiar Directions			
Communicate Understanding	Respond to New Directions			
Asking Questions	Answer Questions	Vocabulary		
Communicate with Others		Number of verbal words		
[Telling About Feeling Sick] ¹		Number of symbols on an assistive device		
[Telling About the Past] ²				
Example of a draft item set for the “Refuse Object” concept on the first version of the ORCA measure:				
4. Please tell us how your child <u>refuses</u> an object like a book, toy, or food.				
		No	Yes, but not consistently	Yes, consistently
4a.	Does your child <u>cry or fuss</u> when they do not want an object?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4b.	Does your child <u>turn their head away or push an object away</u> when they do not want it?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4c.	Does your child <u>kick, grab, throw, or bite</u> when they do not want an object?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4d.	Does your child have <u>a distinct “no” (shaking head, sign/gesture, word/word approximation, or symbol on device)</u> that they use when they do not want an object?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

1. The concept of ‘telling about feeling sick’ was identified for removal based on the psychometric testing results.

2. The concept “telling about the past” is currently included in the ORCA measure but not a part of the scoring metric at this time. In the psychometric study, too few caregivers endorsed this concept for us to feel confident in the stability of the IRT item parameters. This concept represents very high communication ability whose importance was supported by the concept elicitation interviews and cognitive testing, thus, it was not removed. However, more data will need to be collected on the concept before integrating in scoring. When additional data is available, we will be able to add it without disrupting the scoring algorithm for the current set of concepts included in ORCA measure.

Table 2. Demographics and Background Information for Caregiver Participants in the ORCA Measure Development Process.

	Cognitive Testing	Psychometric Testing
	<i>n (%)</i>	<i>n (%)</i>
Total n	24	249
Caregiver Characteristics		
Age, years (<i>Mean / SD</i>)	47.4 / 8.9	41.9 / 8.6
Ethnicity		
Not Hispanic or Latino	23 (96)	228 (92.3)
Hispanic or Latino	1 (4)	172 (6.9)
Race		
White	22 (92)	227 (91.9)
African-American or Black	1 (4)	3 (1.2)
Asian	1 (4)	8 (3.2)
Native American or Alaskan Native	-	1 (0.4)
More than one race	-	4 (1.6)
Relationship to Child		
Mother/Step-Mother	22 (92)	220 (88.4)
Child with AS Characteristics		
Age, years (<i>Mean / SD</i>)	13.8 / 8.2	10.6 / 7.9
Female	14 (58)	112 (45.0)
Ethnicity		
Not Hispanic or Latino	21 (87)	219 (89.8)
Hispanic-Latino	3 (13)	22 (9.0)
Race		
White	21 (88)	215 (86.7)
African-American or Black	1 (4)	3 (1.2)
Asian	-	5 (2.0)
More than one race	2 (8)	22 (8.9)
Genotype		
Deletion Positive	14 (58)	167 (67.1)
Mutation/UBE3A	5 (21)	51 (20.5)
Imprinting (ICD)	2 (8)	10 (4.0)
Uniparental Disomy (UPD)	3 (13)	21 (8.4)

Table 3: ORCA communication ability scores by Angelman syndrome genotype and age cohort.

	n (%)	Mean (STD)	95% Confidence Limits for Mean
AS genotype			
Deletion positive	167 (67.1%)	46.2 (8.4)**	45.0 - 47.5
Mutation of the UBE3A gene	51 (20.5%)	56.6 (7.6)	54.4 - 58.8
UPD	21 (8.4%)	60.2 (6.2)	57.3 - 63.0
Imprinting Center Defect (ICD)	10 (4.0%)	57.8 (5.1)	54.2 - 61.5
Age¹	n (%)	Mean (STD)	Minimum - Maximum
2 to 4 years old	72 (29.0)	44.5 (9.1)	25.8 – 63.6
5 to 12 years old	88 (35.5)	52.7 (8.9)	28.9 – 75.9
13 to 35 years old	88 (35.5)	51.7 (9.0)	31.6 – 73.1

**Mean caregiver scores for individuals with deletion positive AS were significantly different from the other three groups; $p < .001$. ¹One participant did not report age of the child.

Appendix 1. ORCA Item Tracking Matrix from 2 Rounds of Cognitive Testing among Caregivers of Individuals with Angelmen Syndrome

Communication Domain (or #)	Round 1 Wording	Round 2 Wording	Post Round 2 Wording	Reason for Change
Gestures/Signs Definition	When we say gestures or signs, we mean an arm or body movement with specific meaning. Gestures or signs could be formal (like American Sign Language) or informal (a gesture that only your child uses or was modified by your family). Pointing would be considered a gesture.	When we say gestures or signs, we mean an arm or body movement with specific meaning. Gestures or signs could be formal (like American Sign Language) or informal (a gesture that only your child uses or was modified by your family). Eye gaze, pointing and head nodding would be considered gestures.	No further changes	In Round 1, many caregivers asked if eye gaze and head nodding counted as gestures. We added these two behaviors to clarify the meaning of gestures. The re-worded definition worked well in Round 2.
Word/Word Approximation Definition	When we say words or word approximations we mean noises with a specific meaning(s). These can be actual words like “mama” or word approximations like “ba” meaning “bubbles”.	When we say words or word approximations we mean noises with a specific meaning(s). These can be standard words like “mama” that other people can understand, or word approximations that are part of a word like “ba” meaning “bubbles”. Names for people would be considered words or word approximations.	No further changes	In Round 1, participants had a hard time discerning the difference between words and WA. We changed the words definition to clarify we meant understandable or universal words. We also bolded and underlined each term so they could be more easily distinguishable. Many caregivers also asked if names would count as words or WA, so this item was also added to the definition. The modification worked well in Round 2.
Augmentative and Alternative Communication Devices Definition	When we say AAC devices, we mean anything that your child uses to assist in their communication. The device could be high tech (like an application on an iPad or a Dynavox) or low tech (like picture boards).	When we say AAC devices, we mean anything that your child uses to assist in their communication. The device could be high tech (like an application on an iPad or a Dynavox) or low tech (like picture boards). When we say symbols , we mean words, pictures or numbers on your child’s device.	When we say AAC devices, we mean anything that your child uses to assist in his/her communication. The device could be high tech (like an application on an iPad or a Dynavox) or low tech (like picture boards). When we say symbols , we mean words, pictures or numbers on your child’s device that he/she	In Round 1, participants mentioned that we ask about symbols in subsequent questions, but do not define what we mean by “symbols” in the original AAC definition. This wording was added to the definition and worked well in Round 2. However, participants in both rounds were selecting this modality, even if the child was not using the device in a meaningful or intentional

			<i>intentionally</i> selects to communicate something to you.	way. Therefore, we added a note about intentionality to the definition after Round 2.
1A - # of gestures/signs	How many individual gestures/signs does your child use regularly? <input type="checkbox"/> _____	In the past 30 days, how many unique gestures/signs did your child use? <input type="checkbox"/> 1-3 gestures/signs <input type="checkbox"/> 4-6 gestures/signs <input type="checkbox"/> 7-9 gestures/signs <input type="checkbox"/> 10-15 gestures/signs <input type="checkbox"/> 16-20 gestures/signs	In the past 30 days, how many individual gestures/signs did your child use? <input type="checkbox"/> 1-3 gestures/signs <input type="checkbox"/> 4-6 gestures/signs <input type="checkbox"/> 7-9 gestures/signs <input type="checkbox"/> 10-15 gestures/signs <input type="checkbox"/> 16-20 gestures/signs	In Round 1, participants expressed that they would prefer ranges because many did not know the exact number of gestures/signs their child uses (some filled in ranges on the questionnaire itself). Therefore, ranges were added as answer options. We changed “individual” to “unique” in hopes of assisting caregivers in recalling the number of gestures. However, “unique” did not perform well in Round 2, as participants interpreted this word as modified or made up gestures/signs that are unique to the child. The ranges performed well so in the final version we reverted to “individual gestures/signs” while keeping the ranges as the response option.
1B – stringing gestures/signs together	Does your child string gesture/signs together to communicate one message? <input type="checkbox"/> No <input type="checkbox"/> Yes If yes, how many gestures/signs will your child string together to communicate one message? <input type="checkbox"/> _____	In the past 30 days, did your child string gestures/signs together to communicate one message (example: “I want banana”)? <input type="checkbox"/> No <input type="checkbox"/> Yes If yes, on average, how many gestures/signs did your child string together to communicate one message (example: “I want banana” would count as 3 gestures/signs)? <input type="checkbox"/> 2-3 gestures/signs <input type="checkbox"/> 4-5 gestures/signs <input type="checkbox"/> More than 5 gestures/signs	In the past 30 days, did your child put gestures/signs together to communicate one message (example: “I want banana”)? <input type="checkbox"/> No or only once <input type="checkbox"/> Sometimes <input type="checkbox"/> Yes, almost all the time If yes, on average, how many gestures/signs did your child put together to communicate one message (example: “I want banana” would count as 3 gestures/signs)? <input type="checkbox"/> 2-3 gestures/signs	We added an example after Round 1 to clarify the meaning of stringing. Participants were also confused about how to count the number of signs/gestures their child strings together, so we added a note after the example in the second question for participants to reference. The example and counting instructions worked well in Round 2. Participants in Round 1 were thinking about one instance where their child strung together gestures/signs, so “on average” was added to capture only the number

			<input type="checkbox"/> 4-5 gestures/signs <input type="checkbox"/> More than 5 gestures/signs	<p>their child consistently strings together. We also added ranges because participants were providing ranges in the open response option in Round 1. These modifications worked well in Round 2.</p> <p>“Stringing” could not be easily translated according to our translatability review; therefore, we changed stringing to “put together”. This change worked well in the second half of Round 2 interviews. We also changed the response options to the first question after Round 2 to make it consistent with other stringing questions.</p>
1D – pairing sound and gesture together		<p>In the past 30 days, did your child pair a gesture/sign with a sound in order to communicate something to you?</p> <input type="checkbox"/> No or only once <input type="checkbox"/> Sometimes <input type="checkbox"/> Yes, almost all the time	No further changes	<p>Many participants in Round 1 mentioned their child often pairs multiple modalities together to communicate one message. The communication experts on our team stated that pairing a sound with a gesture could represent a more advanced level of communication, therefore we added this question in hopes that it would help detect improvements in communication over time. This question worked well after initial implementation.</p>
# of word approx.	<p>How many word approximations does your child have?</p> <input type="checkbox"/> _____	<p>In the past 30 days, how many unique <u>word approximations</u> did your child use?</p> <input type="checkbox"/> 1-3 word approximations <input type="checkbox"/> 4-6 word approximations <input type="checkbox"/> 7-10 word approximations <input type="checkbox"/> More than 10 word approx.	<p>In the past 30 days, how many individual <u>word approximations</u> did your child use?</p> <input type="checkbox"/> 0 word approx. <input type="checkbox"/> 1-3 word approx. <input type="checkbox"/> 4-6 word approx. <input type="checkbox"/> 7-10 word approx.	<p>In Round 1, participants expressed that they would prefer ranges because many did not know the exact number of WA their child uses (in addition, some filled in ranges on the questionnaire itself). We changed “individual” to “unique” for Round 2 in hopes of assisting</p>

			<input type="checkbox"/> More than 10 word approx.	<p>caregivers in recalling the number of WA their child uses. However, “unique” did not perform well in Round 2, as participants interpreted this word as modified or made up sounds that are unique to the child. The ranges performed well so in the final version we reverted to “individual WA” while keeping the ranges as the response option.</p> <p>A 0 response option was added half way through Round 2 because one participant noted it might be possible that a caregiver could select words/word approximations initially (in question #1), but their child may only use one or the other. This modification worked well for the remainder of Round 2.</p>
# of words	<p>How many words does your child have?</p> <input type="checkbox"/> _____	<p>In the past 30 days, how many unique <u>words</u> did your child use?</p> <input type="checkbox"/> 1-3 words <input type="checkbox"/> 4-6 words <input type="checkbox"/> 7-10 words <input type="checkbox"/> More than 10 words	<p>In the past 30 days, how many individual words did your child use?</p> <input type="checkbox"/> 0 words <input type="checkbox"/> 1-3 words <input type="checkbox"/> 4-6 words <input type="checkbox"/> 7-10 words <input type="checkbox"/> More than 10 words	<p>In Round 1, participants expressed that they would prefer ranges because many did not know the exact number of words their child uses (in addition, some participants filled in ranges on the questionnaire itself). We changed “individual” to “unique” for Round 2 in hopes of assisting caregivers in recalling the number of words their child uses. However, “unique” did not perform well in Round 2, as participants interpreted this word as modified or made up sounds that are unique to the child. The ranges performed well so in the final version we reverted to “individual words” while keeping the ranges as the response option.</p>

				A 0 response option was added half way through Round 2 because one participant noted it might be possible that a caregiver could select words/word approximations initially (in question #1), but their child may only use one or the other. This modification worked well for the remainder of Round 2.
Stringing words or WA together	<p>Does your child string words or word approximations together to communicate one message?</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Yes</p> <p>If yes, how many words or word approximations will your child string together to communicate one message?</p> <p><input type="checkbox"/> _____</p>	<p>In the past 30 days, did your child string words or word approximations together to communicate one message (example: “I want banana”)?</p> <p><input type="checkbox"/> No or only once</p> <p><input type="checkbox"/> Sometimes</p> <p><input type="checkbox"/> Yes, almost all the time</p> <p>If yes, on average, how many words or word approximations did your child string together to communicate one message (example: “I want banana” would count as 3 words/word approximations)?</p> <p><input type="checkbox"/> 2-3 word or word approximations</p> <p><input type="checkbox"/> 4-5 word or word approximations</p> <p><input type="checkbox"/> More than 5 word or word approximations</p>	<p>In the past 30 days, did your child put words or word approximations together to communicate one message (example: “I want banana”)?</p> <p>No changes to second question.</p>	<p>An example was added after Round 1 to clarify the meaning of stringing. Participants were also confused about how to count the number of words/WA their child strings together, so we added a note after the example in the second question for participants to reference. The example and counting instructions worked well in Round 2.</p> <p>Participants in Round 1 were thinking about one instance where their child strung together words/WA, so “on average” was added to capture only the number their child consistently strings together. We also added ranges because participants were providing ranges in the open response option. These modifications worked well in Round 2.</p> <p>“Stringing” could not be easily translated according to our translatability review; therefore, we changed stringing to “put together”. This worked well for the remainder of Round 2.</p>

<p># of AAC symbols</p>	<p>How many symbols does your child use regularly?</p> <p><input type="checkbox"/> _____</p>	<p>In the past 30 days, how many symbols did your child use?</p> <p><input type="checkbox"/> 1-10 symbols</p> <p><input type="checkbox"/> 11-20 symbols</p> <p><input type="checkbox"/> 21-30 symbols</p> <p><input type="checkbox"/> More than 30 symbols</p>	<p>In the past 30 days, how many symbols did your child intentionally use?</p> <p><input type="checkbox"/> 1-10 symbols</p> <p><input type="checkbox"/> 11-20 symbols</p> <p><input type="checkbox"/> 21-30 symbols</p> <p><input type="checkbox"/> More than 30 symbols</p>	<p>In Round 1, participants expressed that they would prefer ranges because many did not know the exact number of symbols their child uses (some participants filled in ranges on the questionnaire itself). Therefore, we added ranges and they worked well in Round 2.</p> <p>In Round 2, some participants were counting symbols that their child randomly/unintentionally uses. Therefore, we added “intentionally” to clarify we want the # of symbols their child uses in a meaningful, consistent way.</p>
<p>Stringing AAC symbols together</p>	<p>Does your child string symbols together to communicate one message?</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Yes</p> <p>If yes, how many symbols will your child string together to communicate one message?</p> <p><input type="checkbox"/> _____</p>	<p>In the past 30 days, did your child string 2 or more symbols together to communicate one message (example: “I want banana” could count as 2 or 3 symbols depending on how your child’s device is set up)?</p> <p><input type="checkbox"/> No or only once</p> <p><input type="checkbox"/> Sometimes</p> <p><input type="checkbox"/> Yes, almost all the time</p> <p>If yes, on average, how many symbols did your child string together to communicate one message?</p> <p><input type="checkbox"/> 2-3 symbols</p> <p><input type="checkbox"/> 4-5 symbols</p> <p><input type="checkbox"/> More than 5 symbols</p>	<p>In the past 30 days, did your child put together 2 or more symbols to communicate one message (example: “I want banana” could count as 2 or 3 symbols depending on how your child’s device is set up)?</p> <p>If yes, on average, how many symbols did your child put together to communicate one message?</p> <p><input type="checkbox"/> 2-3 symbols</p> <p><input type="checkbox"/> 4-5 symbols</p> <p><input type="checkbox"/> More than 5 symbols</p>	<p>We added an example after Round 1 to clarify the meaning of stringing. Participants were also confused about how to count the number of symbols their child strings together, so we added a note after the example for participants to reference. The example and counting instructions worked well in Round 2.</p> <p>Participants in Round 1 were thinking about one instance where their child strung together symbols, so “on average” was added to capture only the number their child consistently strings together. We added ranges because participants were providing ranges in the open response option. The modifications worked well in Round 2.</p>

				<p>“Stringing” could not be easily translated according to our translatability review; therefore, we changed stringing to “put together”. This worked well in the second half of Round 2 interviews.</p>
Requesting AAC device	<p>If their device is not around, does your child let you know that they need their device to tell you something? (could be eye gaze, gesture/sign for device, word/word approximation, etc.)</p> <ul style="list-style-type: none"> <input type="checkbox"/> No <input type="checkbox"/> Yes, but not consistently <input type="checkbox"/> Yes consistently 	<p>In the past 30 days, when their device was not around, did your child let you know that they need their device to tell you something or request something? (Could be eye gaze, gesture/sign for device, word/word approximation, etc.)</p> <ul style="list-style-type: none"> <input type="checkbox"/> No or only once <input type="checkbox"/> Sometimes <input type="checkbox"/> Yes, almost all the time 	<p>In the past 30 days, when his/her device was not around, did your child let you know that he/she needs his/her device to tell you something or request something? (Could be eye gaze, gesture/sign for device, word/word approximation, etc.)</p> <ul style="list-style-type: none"> <input type="checkbox"/> No or only once <input type="checkbox"/> Sometimes <input type="checkbox"/> Yes, almost all the time <input type="checkbox"/> Not applicable 	<p>We added “request something” to add specificity to the question to assist caregivers to recall specific scenarios where their children may have requested their device. Prior to adding this phrase, participants were frequently thinking about their child going to get the device, rather than using another mode of communication to request the device.</p> <p>Not applicable was added after Round 2 because some caregivers stated that their children wear the device via a harness so they would never need to request the device.</p>
Turn Taking		<p>Next we want to know how you have a back and forth conversation with your child. We are interested in the number of turns he/she takes in conversations. The next example shows 2 turns taken: Caregiver: What do you want for breakfast? Child: [Points to cabinet for oatmeal] Caregiver: What do you want in your oatmeal?</p>	<p>Next we want to know how you have a conversation with your child. We are interested in the number of turns he/she takes in conversations. The next example shows 2 turns: Caregiver: What do you want for breakfast? Child: [Points to cabinet for oatmeal] Caregiver: What do you want in your oatmeal?</p>	<p>We added this question after Round 1 per the suggestion of our SLP colleagues. They stated it is an important aspect of pragmatic communication.</p> <p>Changes were made to this question after Round 2 because participants were confused by the phrase “back and forth”, so it was removed. We added a 0 response option because some participants said their child cannot do this activity or have not done it in the past 30 days and we needed to account for this.</p>

		<p>Child: [uses device to say] Strawberries</p> <p>In the past 30 days, about how many conversational turns did your child take during one conversation?</p> <p><input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 or more</p>	<p>Child: [uses device to say] Strawberries</p> <p>In the past 30 days, about how many turns in conversations did your child have during one conversation?</p> <p><input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 or more</p>	<p>Even with the modifications, participants had difficulty with what counted as a communicative act and a turn in conversation. As such, this question will be reevaluated after the cross-sectional study.</p>
Seek Attention	Does your child <u>make a specific noise</u> to seek attention?	Did your child <u>make a specific sound</u> to seek attention?	No further changes	We changed noise to sound per the suggestion of the SLP on our team and to be consistent with our definitions.
Seek Attention	Does your child <u>use a specific gesture/sign</u> to seek attention?	Dropped	No further changes	This item was dropped because many participants ignored the word “specific” in the question. Our intention was for participants to think about dedicated signs to seek attention, however this question did not elicit those responses. Our SLP colleagues also confirmed it is unlikely that children would be using modified gestures or signs to seek attention, as it is not a common way for anyone to seek attention.
Seek Attention	Does your child <u>use a device</u> to seek attention?	Did your child <u>use a specific symbol on a device</u> to seek attention?	No further changes	“Specific” was added to this item after Round 1 because some participants mentioned that their child might press random buttons on their device to get attention. However, we are only interested in times when the child is intentionally selecting a specific symbol that is being consistently used by the child to seek attention (like “look at me!”)

Direct Attention		Did your child <u>use a specific symbol on a device</u> to direct your attention?	Did your child <u>use a symbol on a device</u> to direct your attention?	The term “specific” was dropped from this item after Round 2. Children can direct attention to a multitude of things using a symbol on their device. “Specific” was unnecessary here since most caregivers set up the device for the specific needs of the child and their speech therapy goals.
Direct Attention	Does your child <u>use a gesture/sign and a word (or word approximation) together</u> to direct your attention?	Did your child <u>use a gesture/sign and a sound together</u> to direct your attention?	No further changes	Many participants in Round 1 said their child does not use words/WA but does use gestures and sounds together frequently. After consulting with our SLP colleagues, they also suggested changing words/WA to sounds, as a child who uses words/WA would probably use those alone to direct attention and sounds would be more appropriate here. This suggestion is also reflected in the data we collected.
Direct Attention	This question is asking if your child uses language to direct your attention. This could be in specific ways, like signing "ball" for you to look at the ball. Your child could use gestures/signs, specific vocalizations (i.e. words or word approximations), or symbols on their device.	These questions are asking if your child used language to direct your attention in the past 30 days. This could be in specific ways, like signing "ball" for you to look at the ball.	These questions are asking if your child used language to direct your attention in the past 30 days. This could be in specific ways, like using sign language to say "ball" for you to look at the ball.	After Round 1, we decided to take the modalities out of the instructions because participants said it was repetitive to see the modalities in the directions and the questions. In Round 2, participants were mistaking “signing” to mean pointing. For our purposes, pointing is not considered “labeling”. Therefore, we changed it to “sign language” to be more specific.
Direct Attention	Does your child <u>use a gesture/sign, word or word approximation, or device</u> to let you know specifically what they want you to see?	Did your child <u>use a gesture/sign to label</u> specifically what they wanted you to see?	Did your child <u>use a sign to label</u> specifically what they wanted you to see?	After Round 1, we decided to split this question into 3 questions because participants stated it would represent a significant improvement

		<p>Did your child <u>use a word/word approximation</u> to label specifically what they wanted you to see?</p> <p>Did your <u>child use a symbol on a device</u> to label specifically what they wanted you to see?</p>		<p>in communication ability if their child could do these behaviors with other modalities. Our SLP colleagues also said these separated questions could be a place to show growth in communication ability. This change worked well in Round 2.</p> <p>We also added the term “label” after Round 1 because participants were not understanding that we were looking for the ability to name or label something to direct attention to it; the term “specifically” was being overlooked. Label performed moderately well in Round 2. However, participants in Round 2 were again talking about pointing as “labeling something”. We dropped gesture from the first question to clarify we are only looking for sign language. We will reevaluate these items after the cross sectional study.</p>
Refusal	Does your child <u>turn their head away or push an object away</u> when they do not want it?	Did your child <u>turn their head away, push an object away, or throw an object</u> when they did not want it?	No further changes	Participants in Round 1 thought throwing should be added to this item instead of the kick item. Throwing was not considered a challenging behavior in this population because children often have motor control issues. This change worked well in Round 2.
Request	Does your child have <u>a distinct "yes" (shaking head, sign, or symbol on device)</u> that they use when they want something or are agreeing with you?	Dropped	No further changes	This item was dropped because our SLP colleagues clarified that there are many ways to confirm agreement and it does not necessarily represent an

				improvement in communication ability.
Request	Does your child <u>use a device</u> to ask for an object they want?	Did your child <u>use a symbol on a device</u> to ask for an object they wanted?		“Symbol” was added to be consistent with our definitions and other related items. This modification caused no problems in Round 2.
Request	<p>Think about how your child lets you know that they <u>want something that is not in the same room or that they cannot see.</u></p> <p>Does your child ask for a specific object they want (like a snack, a person, or a toy), <u>even if the object is not around?</u></p>	<p>Think about how your child let you know that they <u>wanted a specific object</u> (like a book, toy or food) <u>that was not in the same room or that they could not see.</u></p> <p>Did your child ask for a specific object they wanted, <u>even if the object was not in sight?</u></p>	<p>Top instructions did not change.</p> <p>Did your child <u>go and get a specific object he/she wanted</u>, even if he/she could not see the object?</p> <p>Did your child <u>use a gesture/sign, word/word approximation, or symbol on a device</u> to ask for a specific object he/she wanted, even if he/she could not see the object?</p>	In Round 1, participants were overlooking the word “specific” and were describing how their child requests things they cannot see, but the caregiver does not always know exactly what the child wants. Caregivers were also often just describing their child getting up and getting what they want, which is an adaptive behavior rather than true communication. We modified this question in Round 2 so that “specific object” would be in the instructions (and underlined). “Not around” was also being misinterpreted as “not in reach”, so this was changed to “not in sight”. Even with our changes in Round 2, these issues still persisted. After Round 2, we decided to create two questions in order to clarify meaning. We will evaluate these items after the cross sectional study.
“More”	Does your child have <u>a specific gesture/sign, word/word approximation, or symbol on their device</u> that they use to ask you for “more” of something?	<p>Did your child <u>use a specific gesture/sign</u> to ask you for “more” of something?</p> <p>Did your child <u>use a word/word approximation, or symbol on a device</u> to ask you for “more” of something?</p>	No further changes	After Round 1, this question was broken into 2 questions to make it consistent with the other item set structures (per our SLP colleagues’ recommendations). Specific gesture/sign became its own item because our colleagues confirmed that this is a lower communicative

				ability than using words or a device. Words and symbols on a device can be universally understood, making those modalities more advanced. This modification caused no problems in Round 2.
Choices	If you give your child options <u>with words only</u> (without showing them choices), does your child choose what they want?	If you gave your child options <u>with words only</u> (without showing them choices), did your child choose what they wanted using a specific gesture/sign, word/word approximation, or symbol on a device?	No further changes	After Round 1, modalities were added to this question because participants were counting when their child “gets up and goes to get the choice” (which we are not counting towards communication ability). This modification worked well in Round 2.
Sick/Hurt		Please tell us how your child <u>let you know he/she was not feeling well, were hurt, or were sick.</u>	Please tell us how your child <u>let you know he/she was not feeling well, were hurt, or were sick.</u> If your child was not hurt or sick in the past 30 days, please select N/A.	N/A response option was added after Round 2 because many participants said their child had not been sick or hurt in the past 30 days. Adding N/A allows us to differentiate between a child who has been ill and unable to do this activity and a child who has not been ill. We will evaluate this modification after the cross sectional study.
Sick/Hurt	Does your child <u>cry, fuss or lay down</u> to tell you that they are not feeling well?	Did your child <u>cry, or fuss</u> to tell you that they were not feeling well?	No further changes.	We dropped “lay down” after Round 1 because after discussing with our SLP colleagues, they said laying down is not necessarily true communication (just a means to an end or self-soothing behavior).
Sick/Hurt	If you ask your child “what hurts”, does your child <u>tell you the name of the body part that hurts?</u>	Did your child use a gesture/sign, word/word approximation, or symbol on a device to <u>name the body part that hurt?</u>	Dropped.	In Round 1, participants were interpreting “tell” as using only verbal words to name the body part. Therefore, we changed “tell” to “name” and added modalities to clarify the meaning. “Name” did not

				work well in Round 2 because caregivers were often just counting pointing as “naming”. Therefore, this item was dropped.
Asking Questions	Does your child <u>use gestures/signs, words/word approximations, or their device</u> to ask you questions?	Did your child <u>use gestures/signs, words/word approximations, or symbols on a device</u> to ask you simple questions (example: Where’s mommy?)? Did your child <u>use gestures/signs, words/word approximations, or symbols on a device</u> to ask you complex questions (example: Why do I have to brush my teeth?)?	No further changes	Participants in Round 1 thought this question was really broad and asked what types of questions we were looking for. In Round 2, we split this question into “simple and complex” questions and added examples so caregivers could see the difference between the two. This change worked well in Round 2.
Past Events	Think about how your child <u>tells you about past events</u> . For example, telling on a sibling or telling you about something that happened at school. Your child might tell you a story through gestures/signs, vocalizations, words or word approximations, or a communication device.	Think about how your child <u>told you about past events</u> . For example, telling on a sibling or telling you about something that happened at school. Your child might tell you a story through gestures/signs, words or word approximations, or symbols on a device.	No further changes	After Round 1, we dropped “vocalizations” to be consistent with our beginning definitions. We also changed “communication device” to “symbols on a device” to be consistent with our beginning definitions and other related items.
Past Events	Does your child <u>tell a story</u> (either true or untrue) that has a logical flow with beginning, middle, and end?	Did your child <u>tell a two-step story about one event that led to another?</u> (example: I ate apple, I threw up) Did your child <u>tell a story</u> (either true or untrue) that had a logical flow with beginning, middle, and end?	Did your child tell a <u>two-part story about one event that led to another?</u> (example: I ate apple, I threw up) Did your child <u>tell a story</u> (either true or untrue) that flowed logically with beginning, middle, and end?	Participants in Round 1 thought this question was really broad and our SLP colleagues upon review also thought this question could be a place to show growth. Therefore, we split this question into two questions for Round 2. This change worked well in Round 2, but some minor word changes were made after Round 2 to clarify meaning.
Greeting	Does your child use gestures/signs to say "hello" or "goodbye" to	Did your child <u>use gestures/signs to say "hello"</u> to people without you telling them to?	No further changes	Some participants in Round 1 said their child could only say hello to people without them telling their

	<p>people without you telling them to?</p> <p>Does your child use words/word approximations, or their device to say “hello” or “goodbye”?</p>	<p>Did your child <u>use words/word approximations or symbols on a device to say “hello”</u> to people without you telling them to?</p> <p>Did your child <u>use gestures/signs to say “goodbye”</u> to people without you telling them to?</p> <p>Did your child <u>use words/word approximations or symbols on a device to say “goodbye”</u> to people without you telling them to?</p>		<p>child to, but their child could not say goodbye. Therefore, we split hello and goodbye greetings into separate questions. We also separated by modality to show growth. This modification worked well in Round 2.</p>
Names for People		<p>Did your child use a specific name, gesture/sign, or symbol on a device <u>for you or another primary caregiver?</u></p> <p>Did your child use specific names, gestures/signs, or symbols on a device <u>for any other person (other than you or another caregiver)?</u></p>	<p>Did your child use a specific name, gesture/sign, word/word approximation or symbol on a device <u>for you or another primary caregiver?</u></p> <p>Did your child use specific names, gestures/signs, words/word approximations or symbols on a device <u>for any other person (other than you or another caregiver)?</u></p>	<p>A couple participants in the beginning of Round 2 asked if we intended to exclude word or WA from these items and did not understand that “use a specific name” meant word. This modification was made half way through Round 2 interviews and worked well.</p>
Self-Reference	<p>Does your child gesture/sign, vocalize or use their device to say <u>“I”, “me”, or their name?</u></p>	<p>Did your child use gesture/sign, word/word approximation or a symbol on a device to say <u>“I”, “me”, or their name in a sentence?</u></p>	<p>No further changes</p>	<p>Some participants in Round 1 thought this question was asking if their child could identify their name or say their name, but we are interested in if a child can refer to themselves in a sentence (i.e. can the child refer to themselves in the first person). Therefore, “in a sentence” was added to make it clearer that we are interested in correct grammatical usage here.</p>

				This modification worked well in Round 2.
Social Games or Activities	<p>Think about any games you and your child might play together. For example, peek-a-boo, chase, or catch.</p> <p>Example: Does your child <u>play</u> games with you or others?</p>	<p>Think about <u>one</u> game or activity your child might have played with you or other people. Games could include peek-a-boo, chase, or catch. Activities could include swimming, playing with Play-Doh, or blocks.</p> <p>Example of change to all items in this domain: Did your child <u>play at least one game or activity</u> with you or others?</p>	No further changes	Multiple participants in Round 1 struggled with this question because they were thinking about multiple games their child plays, and that they may do some of these things for one game only, but not all. Therefore, we added “think about one game or activity” to the directions. Parents of teenagers and young adults had trouble thinking of games they played with their child; therefore, we clarified in the directions that activities would also count. These modifications worked well in Round 2.
Response to Name	Does your child look at you or pause what they are doing <u>when you say their name</u> ?	Excluding times when they ignored you, did your child look at you or pause what they were doing <u>when you said their name</u> ?	No further changes	Multiple participants in Round 1 mentioned that their child does this “sometimes”, but often the child just ignores them (but still understands their name is being called). Therefore, we added a sentence to the beginning of the question to ask parents to disregard times when their child ignores them. This modification worked well in Round 2.
Response to Directions	Please tell us how your child <u>responds to your directions</u> . This first set of questions asks about directions that are part of their <u>daily routine</u> .	Please tell us how your child <u>responded to your directions</u> . This first set of questions asks about directions that are part of their <u>daily routine</u> . Please exclude times when your child may have ignored you.	No further changes	Multiple participants in Round 1 said that their child “sometimes” responds to directions, but often times, the child just ignores them (but still understands the direction). Therefore, we added a sentence to the instructions to ask parents to disregard times when their child

				ignores them. This modification worked well in Round 2.
Response to Directions (Daily Routine)	Does your child respond when you tell them “ <u>no</u> ”? Does your child respond to <u>certain words (stop, bath)</u> when you say them?	Did your child respond when you told them “ <u>no</u> ” or “ <u>stop</u> ”?	No further changes	In Round 1, participants were thinking about stop and bath, not stop <u>or</u> bath like we intended. Many participants also wondered how many one word commands their child has to follow to count as consistently vs. not consistently. This item was subsequently dropped and “stop” was added to the no item because the two commands are similar.
Response to Directions (Daily Routine)	Does your child follow <u>one-step directions</u> that are part of their daily routine? (example: “Get your cup.”)	Did your child follow <u>one-step directions</u> that are part of their daily routine? (example: “Get your cup” or “Let’s take a bath”)	No further changes	As mentioned above, the stop, bath item was dropped and we decided to add these two examples to other items. “Let’s take a bath” was added here because caregivers said this is a very common one step direction they give their child.
Response to Directions (New)	Now think about how your child responds to directions that are <u>new</u> or <u>not</u> part of their daily routine. Example: Does your child follow <u>one-step directions</u> that are new or not part of their daily routine?	Now, think about how your child responded to <u>new directions</u> . Please exclude times when your child may have ignored you. Example of change to all items in this domain: Did your child follow <u>one-step directions</u> that were new ?	No further changes.	Some participants in Round 1 said their child follows directions that are not part of their daily routine, but do not follow new directions. Therefore, we changed this item to reflect just new directions. This modification worked well in Round 2. We also had the same ignoring issue here and added a clause to the instructions asking caregivers to disregard times when their child ignores them. This modification worked well in Round 2.

Isolated Words or Phrases	Does your child show you that they <u>understand a more complex message</u> that they overheard? (example: grandma is visiting next week)	Did your child show you that they <u>understood a more complex message</u> that they overheard? (example: Grandma is bringing cookies)	Dropped.	Participants in Round 1 struggled with the example we provided because they said their child does not understand the concept of time. They may tell their child grandma is visiting, but they were not sure if their child understands “next week”. Therefore, we changed the example to eliminate the time issue. However, in Round 2, some participants were unsure if their child understands both concepts (<u>grandma and cookies</u>), while other participants were overinflating their child’s abilities here. Interviewers know overinflating was occurring because when caregivers were prompted to provide examples of when they have seen this behavior, they could not provide any. Therefore, this item was dropped.
Understanding	<p><u>Are people familiar to your child</u> (parents, teachers, caregivers, siblings) able to understand your child’s communication?</p> <p><u>Are people less familiar to your child</u> able to understand your child’s communication?</p>	<p><u>In general, were people familiar to your child</u> (parents, teachers, caregivers, siblings) able to understand your child’s communication?</p> <p><u>In general, were people less familiar to your child</u> able to understand your child’s communication?</p>	No further changes.	After Round 1, we added “in general” to the beginning of these questions because participants expressed that different people who are familiar to the child understand their child to varying degrees. Therefore, we added “in general” so participants could think about this on average. We also added it to the question below to make it consistent.
Bathroom	<p>Does your child let you know they <u>have gone</u> to the bathroom or that their diaper is dirty?</p> <p>Does your child let you know that they <u>have to go</u> to the bathroom?</p>	Dropped		Questions about bathroom communication were dropped after Round 1 because many participants said they did not know if their child was truly communicating, or if they were just alerting them that they

				were uncomfortable. Also, many different families prioritized different toileting behavior and it was not feasible to account for every scenario in a way that differentiated improvement in communication ability.
--	--	--	--	---

Non-domain changes

	Round 1 Wording	Round 2 Wording	Post Round 2 Wording	Reason for change
Instructions	The goal of this survey is to get a clear understanding about how your child <i>typically</i> communicates. When you are answering the questions, please think about communication skills and behaviors that you have <i>personally observed</i> .	The goal of this survey is to get a clear understanding about how your child <i>typically</i> communicates. When you are answering the questions, please think about communication skills and behaviors that you have <i>personally observed</i> . Please use a ✓ or an X to mark your answers.	The goal of this survey is to clearly understand how your child <i>typically</i> communicates. When you are answering the questions, please think about his/her <i>intentional</i> communication skills and behaviors that you have <i>personally observed and your child initiates</i> . Please use a ✓ or an X to mark your answers. If something is not applicable to your child, please select “No or only once”.	Multiple participants in Round 1 asked “how” they should complete the survey (i.e. filling in bubbles, using a check mark, etc.), so we added this to the instructions to make it consistent. In Round 2, all participants used a ✓ or an X to mark their answers. If something is not applicable to your child, please select “No or only once” was also added to these instructions and the directions before item set #3.
Recall period	The next set of questions are about <u>how</u> your child currently communicates in specific situations.	The next set of questions are about <u>how</u> your child communicated in specific situations over the past 30 days	No further changes	We added a specific recall period because we thought it would help parents focus on consistent communication behaviors they observe (and eliminate communication behaviors caregivers have not seen in a long time that are not consistent or frequent).
Response options	No Yes, but not consistently Yes, consistently	No or only once Sometimes Yes, almost all the time	No further changes	Participants in Round 1 struggled to reliably define the word “consistently”. Some were interpreting it as frequency. Some were interpreting this as the primary modality their child uses. Therefore, more common response options were implemented and worked well in Round 2.

Skip Logic Instructions	Example: If you selected <u>GESTURES/SIGNS</u> , please answer questions 1A and 1B. If not, please skip to 1C.	Example of change to all skip instructions: If you selected <u>GESTURES/SIGNS</u> , please answer questions 1A and 1B. If not, please go to 1C.	No further changes	Our translatability review showed that “skip” was not easily translatable; therefore, we changed this to “go to” based off the company’s recommendation.
Third Person Singular	Example: Did your child <u>cry or fuss</u> when they did not want an object?	Example of change to all items: Did your child <u>cry or fuss</u> when he/she did not want an object?	No further changes	Throughout the survey, we originally referred to the child as them, they or theirs. Our translatability review showed that the third person singular should be used in all verbs in the items when referring to the child. Therefore, all items referring to the child were changed to he/she/him/her. An example of the change is provided.