

Observer-Reported Communication Ability (ORCA) Measure Scoring Manual

(Updated 04-18-2023)

ABOUT THE ORCA MEASURE

The Observer-Reported Communication Ability (ORCA) measure was designed to assess, from the caregiver's perspective, the communication ability of individuals with neurodevelopmental disorders, like Angelman syndrome (AS). Items cover important communication behaviors in numerous areas within expressive, receptive, and pragmatic communication. The content included in the ORCA measure was centered on the aspects of communication that are meaningful to the families of individuals with severe communication deficits, and was informed by in-depth concept elicitation interviews with 22 caregivers of individuals with AS and 6 communication experts. Manuscripts describing the initial development work can be found here:

- <u>https://www.aaidd.org/docs/default-source/prepressarticles/ajidd-d-21-</u>00022 r2.pdf?sfvrsn=1ee63d21 0
- <u>https://www.aaidd.org/docs/default-source/prepressarticles/validation-of-the-observer-reported-communication-ability-measure-for-individuals-with-angelman-</u>syndrome.pdf?sfvrsn=a5323d21 0

The ORCA measure includes 84 total questions, and takes on average about 15 minutes for caregivers to complete. Seventy items ask about observable behaviors within 22 concepts that cover expressive, receptive and pragmatic areas of communication (Table 1). Caregivers indicate if their child has exhibited the behavior in the past 30 days (response options: "no or only once", "sometimes", or "yes, almost all the time"). Fourteen additional items capture information about the individual's unique ways of communicating, including the modalities the individual uses, their current vocabulary, and aspects of language complexity (e.g. how many words/symbols/gestures are used to communicate a single message).

 Table 1: Concepts contributing to estimating communication ability scores on the ORCA

 measure

Expressive	Receptive	Pragmatic
Communication	Communication	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	"Vocabulary"
Communicate Understanding	Respond to New Directions	Number of verbal words
Asking Questions	Answer Questions	Number of symbols on an
Communicate with Others		assistive device
Telling About the Past*		

(c) Copyright 2020. Duke University. All Rights Reserved. Developed by Duke Center for Health Measurement, Duke University School of Medicine. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.



Observer-Reported Communication Ability Measure

*The concept "telling about the past" is currently included in the ORCA measure but is not part of the scoring metric at this time. In the psychometric study, too few caregivers endorsed this concept for the investigators 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, this concept can be added without disrupting the scoring algorithm for the current set of concepts included in the ORCA measure.

RATIONALE FOR SCORING APPROACH

Most multi-item scales (clinical outcome assessments) typically add (or average) responses to all items within the scale (e.g., symptom scales like those for fatigue or depression). This approach treats all items as having equal discrimination ability and doesn't adjust for the difficulty or severity of the item for assessing the outcome of interest. However, this simple aggregation approach was not appropriate for the ORCA measure due to the existing *double hierarchy* in both observable behaviors and communication concepts as identified by the formative concept elicitation work.

There are a number of ways that individuals with communication impairments may communicate within a specific function or for a specific goal. For example, to indicate refusal of an object, an individual may cry or fuss, they may push the object away, or they may communicate 'no' using a distinct 'no' (saying a verbal 'no', shaking their head 'no', or using an AAC device to say 'no'). Crying or fussing represents a lower level of communication ability, while having the ability to indicate a distinct 'no' represents higher levels of ability. In concept elicitation interviews with caregivers of individuals with AS, caregivers often reported that their child phased out or dropped certain behaviors as their skills improved (for example, their child would be less likely to cry or fuss if they could shake their head 'no' or say 'no'). Thus, simply adding up (or averaging) behaviors across items would be inappropriate, as it would weigh skills inappropriately within each concept, and penalize children who phased out lower level skills with the addition of higher level skills. Thus, within a concept, we worked with speech-language pathologists (SLPs) to design a hierarchy of ability levels that reflect developmentally appropriate increases in a child's ability to "master" higher levels of communication ability. (See Appendix #1 for the levels within a concept). Mastery meant that individuals with AS could (and did) perform the behavior frequently and consistently over the past 30 days, and was indicated by caregivers choosing the "Yes, almost all the time" response option.

There also is natural hierarchy in communication at the concept level; different concepts represent higher or lower levels of communication ability. For example, the concept of "Responding to Name" was generally easy for most individuals with AS to master, but the concept of "Asking Questions" represents a higher communication ability level, as very few individuals with AS are able to ask questions consistently and frequently. Thus, any scoring system for the ORCA measure requires the scoring algorithm to be based both on the difficulty of the communication concept and the relationship of the concept to communication ability. As described in the scoring section, item response theory (IRT) scoring is used to adjust for the characteristics of each ORCA concept. The NIH's Patient-Reported Outcomes Measurement Information System (PROMIS) also uses IRT-based scoring for all of their measures.

(c) Copyright 2020. Duke University. All Rights Reserved. Developed by Duke Center for Health Measurement, Duke University School of Medicine. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.



SCORING THE ORCA MEASURE

As described above, the scoring approach for the ORCA measure is different from many other COAs by design, and reflects the unique complexities of communication ability. Thus, the ORCA measure cannot be scored by hand, and an accompanying SAS macro must be used. Please note, the SAS macro can also be run using R. There are 23 concepts in the ORCA measure that are currently included in the assessment of overall communication ability (Table 1). Twenty of the 23 concepts include multiple items from the ORCA measure to evaluate how well an individual masters the behavior (e.g., "seek attention", "respond to name", "greeting"). Three of the 23 concepts only include one item each from the ORCA measure. Appendix #1 provides a list of each of the 23 communication concepts (organized by expressive, receptive, and pragmatic forms of communication) and the representative items and item numbers from the ORCA measure.

For the three concepts that use only a single ORCA item (i.e., "Number of words", "number of symbols", and "turns in conversation"), they have five ordinal response categories with responses to more words, more symbols, or more turns in conversation, respectively, representing higher communication ability. For the other twenty concepts, the ORCA items associated with the concepts have three response options of "No or only once", "Sometimes", and "Yes, almost all the time". In the far right column of the table in Appendix #1, we indicate a hierarchical ordering of communication ability level to differentiate which behaviors represent lower ability (level 0) versus higher ability (higher level numbers) *within each concept*. Each concept can have different ability levels; for example, "Seek attention" has 6 levels (0 to 5) and "Refuse object" has 4 levels (0 to 3). In addition, ability levels are not equivalent across concepts (e.g., a level 3 for the concept of "Greeting" does not equal a level 3 for the concept of "Making Choices"). The ability levels assigned to items within a concept were evidence-based and developed directly from the concept elicitation interviews with caregivers and communication experts. Levels were additionally confirmed by a group of nine SLPs, including two investigators from Duke University.

Based on the caregiver's answers to ORCA items, the individual receives an ability level "score" for each concept (e.g. "seeking attention", "refusal") based on the highest leveled behavior within the concept that the individual had 'mastered'. Mastery meant that individuals with AS could (and did) perform the behavior frequently and consistently over the past 30 days, and was indicated by caregivers choosing the "Yes, almost all the time" response option.

Samejima's graded item response theory (IRT) model was used to adjust for the differing ability levels among the concepts (and levels within a concept) by fitting item parameters at the concept level. IRT PRO (version 4.2) was used to estimate the parameters using marginal maximum likelihood estimation.

The sample used to calibrate the ORCA concepts were caregivers of individuals with AS. Based on responses from 249 caregivers, investigators estimated the IRT model parameters for the 23 concepts with the parameters provided in Appendix #2. The "*a*" parameter indicates how well

(c) Copyright 2020. Duke University. All Rights Reserved. Developed by Duke Center for Health Measurement, Duke University School of Medicine. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.



Observer-Reported Communication Ability Measure

the concept discriminates among individuals with different levels of communication ability. The "b" parameters indicate the levels of communication ability that are assessed by each concept. In other words, the IRT "b" parameters adjust for the "difficulty" of each ORCA concept, with more "difficult" concepts (larger "b" parameters) representing higher-level communication skills. The investigators also used the IRT model to estimate the ORCA measure total score for each individual. Based on a caregiver's response to the 23 communication concepts included in the ORCA measure, an IRT score based on the expected *a posteriori* (EAP) algorithm was estimated. The EAP scores were standardized to a mean of 0 and standard deviation of 1 in the calibration sample of individuals with AS. The EAP scores were then transformed to a T-score metric with a mean of 50 and standard deviation of 10.

To summarize, the ORCA measure is scored using an IRT-based model to reflect the caregiver's observations of an individual's communication ability over the past 30 days via reported mastery of specific behaviors within 23 concepts reflecting expressive, receptive, and pragmatic types of communication.

INTERPRETING ORCA SCORES

The ORCA measure currently produces a single score that is an estimate of an individual's overall level of communication ability. The ORCA T-scores were standardized based on the original calibration sample of individuals with AS to have a mean of 50 and standard deviation of 10. Higher ORCA T-scores reflect greater communication ability; the mastery of expressive, receptive, and pragmatic types of communication and higher vocabularies for verbal words and symbols on assistive devices. The ORCA T-score range is from 25.8 to 83.8. Figure 1 shows the ORCA T-score distribution for the original calibration sample.



Figure 1 Distribution of ORCA Scores for Sample of 249 Caregivers of Children with AS

(c) Copyright 2020. Duke University. All Rights Reserved. Developed by Duke Center for Health Measurement, Duke University School of Medicine. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.



Observer-Reported Communication Ability Measure

The investigators have looked into creating sub-scores of communication by expressive, receptive, and pragmatic forms of communication, but the empirical evidence (from the original calibration sample) supports that there is an overall general factor that accounts for a majority of variance observed in the item response data. If scores are produced for each form (expressive, receptive, pragmatic), they are highly correlated. Options for scoring at the expressive, receptive, and pragmatic levels will be explored as more data is collected.

Currently, an exploratory analysis of possible minimal clinically important differences (MCID) estimates has been performed using the cross-sectional calibration sample. Using a distributionbased approach, a Cohen's effect size between .2 (small) and .5 (moderate) was targeted. This means that roughly a 2 to 5 point difference (i.e., .2 to .5 SDs) on the ORCA T-score metric would represent a MCID. Differences in scores between individuals with deletion positive AS versus other genotypes was explored, as there are reports that individuals with deletion positive genotype have a more severe phenotype than those with other mutations. Differences in scores between individuals with deletion positive AS and the other three genotypes (*mean* $\Delta = 9.98$ -13.97), far exceeded this threshold. The investigators recognize that a stronger approach to establishing meaningful differences would be based on clinical anchors, and this work is ongoing.

MISSING DATA RECOMMENDATIONS

If scores are produced, the scoring algorithm assumes the data is missing at random. The more missing responses on the ORCA measure, the less confidence (higher standard error) we would have in the estimates of communication ability. As a rule of thumb, it's recommended that if 50% or more of the total items (42/84 items) on the ORCA measure are missing or skipped, the ORCA score should not be estimated for that individual. For some contexts of use, a higher threshold for amount of missingness should be considered.

In addition, the SAS macro is set up so that if any item within a concept is missing, then that concept is set to missing within the scoring algorithm. For example, the "Direct Attention" concept has 9 items that map to 5 mastery levels of communication ability. If one or more of the nine items are missing, then the "Direct Attention" concept is set to missing and the ORCA score is estimated based on responses to the other 22 concepts included in the ORCA measure. If your team prefers a different decision rule, please reach out to the investigators to discuss options.

INSTRUCTIONS FOR SCORING WITH R

Start by preparing the item response data by converting the raw item-level responses to the concept-level scores using the algorithm specified in Appendix 1. This step has been programmed in the SAS macro. The SAS code maybe helpful to be translated into R syntax. The SAS macro, "ScoreORCA v2.1.sas", can be opened by Notepads.

Please also refer to the "Item and concept scoring algorithm.xlsx" spreadsheet. Read the conceptlevel response data into R to create the response dataset "resp.data". Please name the columns using the "ItemID" given in the algorithm Excel document above.

⁽c) Copyright 2020. Duke University. All Rights Reserved. Developed by Duke Center for Health Measurement, Duke University School of Medicine. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.



Next, read the IRT parameter file "ItemIRTParametersV2.xlsx" into R to create the item parameter dataset "ipar". Then, define an out file 'sfile' that you want to save the ORCA T scores and SEs. Include the attached calcEAP R syntax and call function thetaSE.eap on (ipar, resp.data)

Help document:

The R function thetaSE.eap was written based on the R function eap(). R documentation about the function can be found here -

https://www.rdocumentation.org/packages/irtoys/versions/0.2.2/topics/eap.