Duke University School of Medicine

Observer-Reported Communication Ability (ORCA) Measure Scoring Manual

Version date: February 4, 2025

Feb 4, 2025

Contents

1. UPDATES TO THE SCORING MANUAL

This manual has been updated from its original version (dated 04-18-2023). The scoring algorithms discussed in this manual apply to ORCA versions 1.0 (Angelman Syndrome) and 2.0 (Rett Syndrome).

The previous version of this manual only discussed scoring at the "Mastery" level for the ORCA measure, or the extent to which the caregiver reported that their child performed a communication behavior "almost all the time" over the past 30 days. While the "Mastery" level option is still available, scoring at the "Emerging & Mastery" level has now been added to capture communication behaviors that a child performed "sometimes" over the past 30 days.

Additionally, the previous version of this manual only discussed scoring at the "Overall" communication score level, where a child would receive one single score across expressive, receptive, and pragmatic forms of communication. This manual has been updated with instructions for how to derive individual scores for each form of communication (expressive, receptive, and pragmatic).

For questions, concerns, or feedback, please reach out to Nicole Lucas at nicole.lucas@duke.edu.

2. 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 (NDDs), like Angelman syndrome (AS). Items cover important communication behaviors in numerous areas within expressive, receptive, and pragmatic (social) 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 initially by in-depth concept elicitation interviews with 22 caregivers of individuals with AS and six communication experts. Manuscripts describing the initial development work, along with subsequent work in Rett syndrome, can be found here:

- https://doi.org/10.1352/1944-7558-128.3.185
- https://doi.org/10.1352/1944-7558-128.3.204
- https://doi.org/10.1016/j.ejpn.2023.07.007

The ORCA measure includes 81 total questions and takes, on average, about 15 minutes for caregivers to complete. Seventy items ask about observable behaviors within 23 concepts that cover expressive, receptive, and pragmatic forms 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"). Eleven 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 by form of communication

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 (others & self)
Request Object Out of View	Make Choices	
Request "More"	Respond to Familiar Directions	
Communicate Understanding	Respond to New Directions	
Asking Questions	Answer Questions	
Communicate with Others		
Telling About the Past*		
"Vocabulary"		
Number of verbal words		
Number of symbols on an assistive device		

*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 initial psychometric study, too few caregivers of individuals with AS 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 it into 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.

3. RATIONALE FOR SCORING APPROACH – HIERARCHY OF COMMUNICATION ABILITY

Most multi-item scales (clinical outcome assessments (COAs)) 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 does not adjust for the difficulty or severity of the item for assessing the outcome of interest. However, this simple aggregation approach is 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*, *push the object away*, or communicate refusal using a distinct *'no'* (e.g., saying a verbal '*no*', shaking their head '*no*', or using an AAC device to say '*no*'). Based on input from communication experts, crying or fussing represents a lower level of communication ability, while having the ability to indicate a distinct '*no*' represents higher levels of communication ability. In concept elicitation interviews with caregivers of individuals with AS and other NDDs, 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 communication concept, and penalize children who phased out lower-level skills with the addition of higher-level skills. Thus, within a communication 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 perform higher levels of communication ability.

Figure 1: Questions related to the communication concept of "refusal of an object"

5. Please tell us how your child refused an object like a book, toy, or food.

	In the past 30 days,	No or only once	Sometimes	Yes, almost all the time
5a.	Did your child cry or fuss when he/she did not want an object?	○	○	○
5b.	Did your child turn their head away, push an object away, or throw an object when he/she did not want it?	○	○	○
5c.	Did your child kick, grab, or bite when he/she did not want an object?	○	○	○
5d.	Did your child use a distinct "no" (shaking head, gesture/sign, word/word approximation, or symbol on a device) when he/she did not want an object?	○	○	○

There is also a 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" is generally easier for most individuals with NDDs to master, but the concept of "Asking Questions" represents a higher communication ability level, as very few individuals with AS and other NDDs 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 section #6, item response theory (IRT) scoring is used to adjust for the characteristics of each ORCA communication concept. The NIH's Patient-Reported Outcomes Measurement Information System® (PROMIS®), and communication measures like the Bayley Scales of Infant Development and Vineland Adaptive Behavior Scales, also use IRT-based scoring.

4. OVERALL COMMUNICATION AND COMMUNICATION SUBTYPE (EXPRESSIVE, RECEPTIVE, AND PRAGMATIC) SCORING

The original ORCA scoring algorithm provided a single score that reflected overall communication ability that included aspects of expressive, receptive, and pragmatic forms of communication. Conceptually, these forms are related in that the same modalities (e.g., gestures, signs, vocalizations) are needed to perform the communication concepts under each form. In addition, the original evaluation of the factor structure in the initial study among individuals with Angelman syndrome heavily supported a single dimension of overall communication (see:

https://doi.org/10.1352/1944-7558-128.3.204). The ORCA "overall" score reflected the concepts shown in Table 1, except for the concept of "telling about the past."

Based on end-user feedback, scoring both for overall communication and for expressive, receptive, and pragmatic forms of communication is now available. See Table 1 for concepts included for each form. The end-user should select which option is more suited for their research study. Note that scores are not linked across forms. For example, an individual with scores of 45 on expressive communication and 40 on receptive communication cannot be described as having a higher level of expressive than receptive communication abilities.

5. EMERGING & MASTERY SCORING OPTIONS

The majority of the items included in the ORCA measure have three response options: "No or only once," "Sometimes," and "Yes, almost all the time" (see Figure 1 for example). All items use a 30-day recall period. When an individual performs a communication behavior "sometimes" from the caregiver's observations, this is referred to in the ORCA scoring as "emerging," meaning that the child performs the communication behavior a few times (i.e., more than once), but not "almost all the time." When an individual performs a communication behavior "almost all the time" from the caregiver's observations, this is referred to in the ORCA scoring as "mastery", meaning that the child performed the behavior frequently and consistently over the past 30 days.

As noted earlier, the development team used a well-established and widely-used methodology (IRT) to estimate the ORCA scores reflecting the natural hierarchy of communication ability (see section #2). The IRT model requires large sample sizes to estimate the parameters for the communication concepts and to feel confident about the communication scores. The rare disease setting presents a challenge when applying IRT modeling, as sample sizes may be small. The development team designed the initial scoring algorithm that reflected feedback from communication experts (e.g., speech-language pathologists) and what was feasible from the collected response data in Angelman syndrome. For example, communication experts reviewed the items included for "refusal of an object" (Figure 1) and indicated which behaviors (and response options) may represent ordered levels of hierarchy of ability to refuse an object. At times, levels were merged when sample sizes were too small (or not available) for specific response options. Any merging of levels was confirmed with communication experts.

Originally, the ORCA team released a scoring algorithm that only provided "Mastery" scoring, reflecting when an individual performed a communication behavior frequently or consistently over the past 30 days. Based on end-user feedback, the ORCA team tested and designed a second set of scoring algorithms that captures "Emerging and Mastery" levels of communication ability. Both types of scoring are available for the end user based on their needs.

"Mastery" Scoring Algorithms – This option should be selected if the end-user wants to capture only communication behaviors that are performed consistently (i.e., almost all the time) and not once or sometimes over the past 30 days. Scores for this algorithm will increase over time when the individual performs a higher level of communication behavior (compared to baseline) almost all the time. For example (from Figure 1), when a child at baseline consistently "cries or fusses" to refuse an object (question 5a) to when a child at follow-up consistently has a distinct no (question 5d). The strength of this option is that the algorithm only picks up

consistent behaviors and not behaviors that may have happened infrequently over the course of 30 days. The limitation of this option is the inability to pick up on emerging behaviors that may translate at later time points to more consistent behaviors. The algorithm may also miss out on certain concepts where there is a variety of ways an individual performs a behavior; for example, an individual may greet someone sometimes with a gesture and sometimes with a verbal phrase (e.g., "hello"). In this example, the child consistently greets others but the caregiver reports "sometimes" for different items reflecting different behaviors to greet others.

"**Emerging & Mastery" Scoring Algorithms** – This option should be selected if the end-user wants to capture both emerging ("sometimes") and mastery ("almost all the time") communication behaviors. A limitation is that in order to accommodate both emerging and mastery behaviors into the algorithm, levels that were once separated under the "Mastery" algorithm were merged to accommodate the small sample sizes in the calibration dataset. While this was not desirable, communication experts still agreed to the merging, given the available options. The development team also conducted sensitivity analyses to ensure scores still represented the construct of interest (communication ability). Sensitivity analyses included (but were not limited to) a comparison of raw sum scores to IRT-based scores and the correlation of ORCA scores with scores from other communication measures.

The end user should select the option that best suits their study's goals. The ORCA development team recommends using the "Emerging & Mastery" scoring algorithm. A scientific manuscript is being drafted to describe the scoring algorithms in more detail. This manual will be updated when the article is published.

6. SCORING THE ORCA MEASURE

As described previously, the scoring approach for the ORCA measure is designed to reflect the unique complexities of communication ability, offering end-users flexibility in assessment, and is different than many other COAs by design. Thus, the ORCA measure cannot be scored by hand, and an accompanying SAS macro or R-based program must be used (see Sections 9 and 10). There are 23 concepts in the ORCA measure that are currently included in the assessment of overall communication ability (see Table 1). Sixteen of the 23 concepts include multiple items from the ORCA measure to evaluate how an individual performs the communication concept (e.g., "refuse object," "seek attention," "respond to name," "greeting"). Seven of the 23 concepts only include one item each from the ORCA measure, which is "Number of Words," "Number of Symbols," "Turns in Conversation," "Responding to Name," "Understanding Mood," "Understanding Isolated Words," and "Comfort Others."

The majority of ORCA items have three response options: "No or only once," "Sometimes," and "Yes, almost all the time." The only exceptions are the following single-item concepts: "Number of Words," "Number of Symbols," "Turns in Conversation," "Responding to Name," "Understanding Mood," "Understanding Isolated Words," and "Comfort Others."

For the ORCA "Mastery" Scoring Algorithms, Appendix #1 lists 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 ORCA "Emerging & Mastery" Scoring Algorithms, Appendix #2 should be used and includes 24 communication concepts (separating "use of names" to differentiate their own name and others).

In the far-right column of the tables in Appendices #1 and #2, we indicate a hierarchical ordering of communication ability levels to differentiate which behaviors represent lower ability (level 0) versus higher ability (higher level numbers) *within each communication concept*. Each concept can have different ability levels; for example, in Appendix #1 (Mastery scoring), "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 speech language pathologists, including two investigators from Duke University. Also, sample size was an issue and required merging some levels for the "Emerging and Mastery" algorithm.

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 performed.

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 #3 (for Mastery scoring algorithms) and in Appendix #4 (for the 24 concepts in "Emerging & Mastery" scoring algorithms). The "*a*" parameter indicates how well 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's score for each individual. Based on a caregiver's response to the 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 a 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 behaviors within concepts reflecting expressive, receptive, and pragmatic types of communication.

7. INTERPRETING ORCA T-SCORES

Table 2 provides a list of the available scoring options for the ORCA measure and the range of possible ORCA T-scores.

Table 2. Range of ORCA T-Scores Across Scoring Algorithms

	"Mastery" Scoring Algorithm	"Emerging & Mastery" Scoring Algorithm
	Minimum – Maximum T-scores	Minimum – Maximum T-scores
Overall	25.8 – 83.8	18.0 – 81.2
Expressive	34.6 – 75.9	25.9 – 79.9
Receptive	26.9 – 69.1	22.5 – 71.6
Pragmatic	33.6 – 72.6	29.5 – 68.7

The ORCA T-scores were standardized based on the original calibration sample of individuals with AS to have a mean of 50 and a standard deviation of 10. Higher ORCA T-scores reflect greater communication ability. For example, an individual with an ORCA T-score of 65 would be 1.5 standard deviations above the reference group mean of 50 (the reference group is the original calibration sample of data provided by 249 caregivers of individuals with AS). Note that scores from one algorithm cannot be compared to scores from another algorithm (e.g., comparing an Expressive T-score with a Receptive T-score). They are not linked together.

The ORCA development team is in the process of estimating meaningful score differences (MSDs) and meaningful score ranges (MSRs) consistent with draft FDA guidelines for Patient-Focused Drug Development (Guidance 4). Using a distribution-based 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 distribution-based MSD. The investigators recognize that a stronger approach to establishing MSDs and MSRs would be based on clinical anchors, and this work is ongoing.

8. 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, a two-step approach is recommended.

Step 1 - within a concept, if any item is missing, then the concept score is missing.

Step 2 - if 50% or more of the total concepts (12 out of 23 or 24) on the ORCA measure are missing, the ORCA score should not be estimated for that individual.

This missing data handling approach applies to both the SAS macro and R programs. For some contexts of use, a higher threshold for amount of missingness should be considered. If your team prefers a different decision rule, please reach out to the investigators to discuss options.

9. INSTRUCTIONS FOR SCORING IN SAS

To get ORCA Mastery T-Scores

- Materials
  - SAS/IML software
  - SAS macros: *ScoreORCA_MasteryOverall.sas* to get the overall ORCA T score, *ScoreORCA_MasterySubForms.sas* to get the ORCA Expressive, Receptive, and Pragmatic T scores; ORCA item to concept mastery scoring algorithm has been programmed in the SAS macros.
- Steps
  - Read and understand the instructions in the comment section on top of the SAS code in the SAS scoring macro to be used;
  - Verify IML is enabled in the SAS software by running the example included in the SAS macro programs;
  - Follow the instructions to plug in parameters and call the scoring macro. When the program finishes running, the output ORCA score dataset will appear in the user-specified data directory.

To get ORCA Emerging & Mastery T-Scores

- Materials
  - SAS/IML software
  - SAS macros: *ScoreORCA_EmergingMasteryOverall.sas* to get the overall ORCA T score, *ScoreORCA_EmergingMasterySubForms.sas* to get the ORCA Expressive, Receptive, and Pragmatic T scores; ORCA item to concept emerging-mastery scoring algorithm has been programmed in the SAS macros.
- Steps
  - Read and understand the instructions in the comment section on top of the SAS code in the SAS scoring macro to be used;
  - Verify IML is enabled in the SAS software by running the example included in the SAS macro programs;
  - Follow the instructions to plug in parameters and call the scoring macro. When the program finishes running, the output ORCA score dataset will appear in the user-specified data directory.

10. INSTRUCTIONS FOR SCORING IN R

To ensure the code is easy to use and understand, we provide two formats:

1) Traditional R Script (.R): A simple text file with R code.
2) R Markdown (.Rmd): Combines R code, results, and explanations in one document. "Knitting" the file generates a clean, shareable HTML report.

We recommend using the *R Markdown* file for users less familiar with R code.

To get ORCA Mastery T-Scores

- Materials
  - R software available for download at https://cran.r-project.org/
  - R program: *ScoreORCA_Mastery.R*;

- 'tidyverse' package: You can install by running the code install.packages("tidyverse").
- Concept-level IRT parameter datasets:
  - *IRTParameters_M.csv*;
  - *IRTParameters_M_Expressive.csv;*
  - *IRTParameters_M_Receptive.csv*;
  - *IRTParameters_M_Pragmatic.csv;*
- Example ORCA survey response data for the demo: *ORCAItemResponseSampleData10.csv*

- Steps: **R Script**
  - Read and understand the instructions on top of the R code.
  - Follow the instructions to plug in parameters wherever "user-supplied" appears.
  - Run the R programs for scoring the sample data to verify the work directory and the required R libraries have been installed.
    - Run getwd() if you're unsure of your current working directory.
  - Prepare the ORCA survey item response data to be scored
    - Make sure the item response variables are numeric and have no out-of-range values.
  - Replace the sample data with the actual ORCA response dataset to be scored and rerun the R program.
  - Check the output datasets. You should have a CSV file for each ORCA T score: Overall, Expressive, Receptive, and Pragmatic.

- Steps: **R Markdown**

1. Unzip and Open the Folder:
   - You will see three sub-folders: **Data**, **Output**, and **Parameters**.
     - **Data**: Contains sample data. To run a test, replace User-supplied-data.csv with ORCAItemResponseSampleData10.csv.
       1. **Add your dataset to this folder.**
     - **Parameters**: Contains the Mastery parameter datasets needed for the code.
     - **Output**: Will contain the results, including thetaSE datasets and four new datasets with your original data plus columns for:
       1. Mastery Overall
       2. Mastery Expressive
       3. Mastery Receptive
       4. Mastery Pragmatic
2. Open the R Project:
   - Double-click the .Rproj file in the main folder to open the project.
3. Open the R Markdown File:
   - Once the R Project is open, locate and open the R Markdown file (.Rmd).
4. Update the Dataset Name:
   - Replace "User-supplied-data" with the name of your dataset.
5. Knit the File:
   - In the top menu bar, click **Knit** to generate the output.

Your results will be saved in the **Output** folder.

To get ORCA Emerging & Mastery T-Scores

- Materials
  - R software available for download at https://cran.r-project.org/
  - R program: *ScoreORCA_EmergingMastery.R*;
  - Concept-level IRT parameter datasets:
    - *IRTParameters_EM.csv*;
    - *IRTParameters_EM_Expressive.csv;*
    - *IRTParameters_EM_Receptive.csv*;
    - *IRTParameters_EM_Pragmatic.csv;*
  - Example ORCA survey response data for the demo: *ORCAItemResponseSampleData10.csv*
- Steps
  - Read and understand the instructions on top of the R code.
  - Follow the instructions to plug in parameters wherever "user-supplied" appears.
  - Run the R programs for scoring the sample data to verify the work directory and the required R libraries have been installed.
    - Run getwd() if you're unsure of your current working directory.
  - Prepare the ORCA survey item response data to be scored
    - Make sure the item response variables are numeric and have no out-of-range values.
  - Replace the sample data with the actual ORCA response dataset to be scored and rerun the R program.
  - Check the output datasets. You should have a CSV file for each ORCA Emerging-Mastery T score: Overall, Expressive, Receptive, and Pragmatic.

- Steps: **R Markdown**
  1. Unzip and Open the Folder:
     - You will see three sub-folders: *Data*, *Output*, and *Parameters*.
       - *Data*: Contains sample data. To run a test, replace User-supplied-data.csv with ORCAItemResponseSampleData10.csv.
         1. **Add your dataset to this folder.**
       - *Parameters*: Contains the Mastery parameter datasets needed for the code.
       - *Output*: Will contain the results, including thetaSE datasets and four new datasets with your original data plus columns for:
         1. EM Overall
         2. EM Expressive
         3. EM Receptive
         4. EM Pragmatic
  2. Open the R Project:
     - Double-click the .Rproj file in the main folder to open the project.
  3. Open the R Markdown File:
     - Once the R Project is open, locate and open the R Markdown file (.Rmd).
  4. Update the Dataset Name:

- Replace “User-supplied-data” with the name of your dataset.

5. Knit the File:
   - In the top menu bar, click **Knit** to generate the output.

Your results will be saved in the **Output** folder.

R 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.