Published in final edited form as:

Speech Lang Hear. 2017 ; 20(3): 130–143. doi:10.1080/2050571X.2016.1243747.

Using Rasch Analysis to Examine the Item-Level Psychometrics of the *Infant-Toddler Meaningful Auditory Integration Scales*

Brittan A. Barker, Ph.D.,

Department of Communication Disorders, Louisiana State University, Baton Rouge, LA USA,
ORCID iD 0000-0001-9327-7057

Neila J. Donovan, Ph.D.,

Department of Communication Disorders, Louisiana State University, Baton Rouge, LA USA,
ORCID iD 0000-0002-7235-8357

Anne D. Schubert, M.A., and

Department of Communication Disorders, Louisiana State University, Baton Rouge, LA USA

Elizabeth A. Walker, Ph.D.

Department of Communication Disorders, University of Iowa, Iowa City, IA USA

Abstract

The Infant-Toddler Meaningful Auditory Integration Scales (IT-MAIS; Zimmerman-Phillips et al. 2001) is a popular assessment designed to measure listening skills in children with hearing loss aged 0–3 years. For this study we analyzed the item-level psychometric properties of the IT-MAIS via Rasch analysis to gain further understanding about its validity and reliability. We chose to analyze the psychometric properties of the IT-MAIS because very little information exists regarding its development and validation, although it is widely used to assess listening skills in children with SNHL ages 0 to 3 years pre- and post-CI. Our results indicated that the IT-MAIS items demonstrated less than ideal psychometric properties and the IT-MAIS item order did not reflect the order in which children are expected to develop functional listening skills. Our findings suggest that there is a pressing need for further discussion among researchers and clinicians about 1) how the IT-MAIS is used, and 2) what other valid and reliable outcome measures could be used alongside, or in place of, the IT-MAIS to determine CI candidacy, establish treatment goals, or track progress in listening development in very young children with hearing loss.

Keywords

validity; reliability; Rasch analysis; children; cochlear implants; *Infant-Toddler Meaningful Auditory Integration Scale*; infants; toddlers; hearing loss

Brittan A. Barker is the corresponding author. She is now affiliated with Utah State University, Department of Communicative Disorders and Deaf Education, 2620 Old Main Hill, Logan, UT 84322-2620, USA. Phone: 435.797.0434 brittan.barker@usu.edu.

Portions of this work were presented under the title, “An examination of the validity and reliability of the Infant-Toddler Meaningful Auditory Integration Scales” at The Hearing Across the Lifespan (HEAL) Conference, Cernobbio, Lake Como, Italy in June 2014; at the American Auditory Society Annual Meeting held in Scottsdale, AZ in March 2013; and at the American Cochlear Implant Alliance’s 2013 Symposium held in Washington D.C. in October 2013.

The pediatric cochlear implant (CI) candidacy evaluation for very young children includes a battery of testing to ensure medical and audiometric suitability. The *Infant-Toddler Meaningful Auditory Integration Scale (IT-MAIS)* (Zimmerman-Phillips *et al.* 2001) is a caregiver-report tool often included in this battery. Specifically, the *IT-MAIS* is used to assess and monitor functional listening pre- and post-CI in children aged 0 to 3 years with sensorineural hearing loss (SNHL). Despite the fact that the *IT-MAIS* is the most frequently administered caregiver-report questionnaire by pediatric CI professionals in the United States (U.S.; Uhler and Gifford 2014), little research to date examined its validity and reliability (Zheng *et al.* 2009, Zimmerman-Phillips *et al.* 2000). As our profession advances, it has become increasingly important that researchers and clinicians are knowledgeable consumers who can select the most appropriate assessment to inform their evidence-based practice. Part of this knowledge comes from determining not only how an assessment was developed, but also its validity and reliability (Dollaghan 2007). Based on the limits of the existing literature, we propose it is important to further explore the psychometric properties of the *IT-MAIS* and provide researchers and clinicians with additional information when making decisions about using the *IT-MAIS* in their research and clinical practice.

The *IT-MAIS*

Across the globe, a number of pediatric CI research and clinical programs use the *IT-MAIS* to help determine CI candidacy and track listening development post-implantation in children with SNHL (e.g., Barker *et al.* 2011, Cardon and Sharma 2013). The *IT-MAIS* includes 10 questions developed to measure a child's ability to vocalize, alert to sounds, and derive meaning from sound (Zimmerman-Phillips *et al.* 2001). Via interview format, an experienced pediatric audiologist elicits responses from a parent/guardian about aspects of their child's auditory development. The *IT-MAIS* instructions encourage the administering audiologist to use a "flexible" interview format to elicit optimal responses. The caregiver's responses are then rated on a 0 – 4 Likert scale reflecting the frequency of the child's behaviors (0/*never*, 1/*rarely*, 2/*occasionally*, 3/*frequently*, 4/*always*). The *IT-MAIS* is a criterion-referenced assessment—an assessment that provides a basis for determining an individual's skill level relative to a theoretically motivated and operationally defined domain of content.

The *IT-MAIS* was derived from the *Meaningful Auditory Integration Scales (MAIS)* (Robbins *et al.* 1991) designed to evaluate meaningful use of sound in children with profound SNHL aged 5 years and older. Zimmerman and colleagues (2000) revised the *MAIS* for children aged 0 to 3 years old. During the revision, they replaced 2 *MAIS* items related to device bonding with 2 items related to vocalizations, behavioral markers often associated with listening development in children with CIs (Ertmer and Jung 2012). The authors retained the remaining 8 *MAIS* items originally identified as skills demonstrated by children 5 years and older and included them in the *IT-MAIS*.

IT-MAIS standardization

After development of the *IT-MAIS*, Zimmerman-Phillips and colleagues (2000) validated the tool via a study that assessed the overall scores of 9 children, 18 to 23 months old, based

on *IT-MAIS* scores obtained during pre-CI, hearing-aid trials, and again at 3 months post-CI. Pre-CI the majority of children received scores of *0/never* on all items, indicating that the caregivers *never* witnessed any of the listening or vocalization behaviors. At 3 months post-CI, all caregivers reported an increase in frequency for at least 7 out of 10 items' behaviors. The researchers concluded that the increases noted from pre- to post-CI demonstrated that the *IT-MAIS* was a valuable tool to measure CI candidacy and benefit. However, their results should be interpreted with caution due to the small sample size (N = 9), which may reflect inadequate representation of the pediatric population (age 0–3 years) with severe-profound SNHL. Furthermore, considering the 3-month time lapse between pre-CI and post-CI assessments in their study, there is no way to know from their data whether the increased frequency of behaviors represented post-CI benefit or typical cognitive and physical development.

Since Zimmerman-Phillips and colleagues' (2000) validation, no one to date has further explored the validity or reliability of the *IT-MAIS*. However, two studies did attempt to develop norms for the *IT-MAIS* based on performance of children with normal hearing (Kishon-Rabin *et al.* 2001, Zheng *et al.* 2009). Kishon-Rabin and colleagues established developmental norms for the *IT-MAIS* based on evaluations of 109 Hebrew- and Arabic-speaking children with normal hearing while Zheng and colleagues (2009) administered an *IT-MAIS* translated into Mandarin-Chinese to 120 Chinese children with normal hearing thresholds, and with native Mandarin-speaking caregivers. Although their sample sizes are remarkable, and the results shed light on listening development in infants and toddlers with normal-hearing thresholds, it is difficult to generalize these data to infants and toddlers with severe to profound SNHL who are brought up in families utilizing spoken English.

The *IT-MAIS*, nonetheless, appears to have face validity since it is widely used to establish treatment goals and track listening development in pediatric CI users over time (Uhler and Gifford 2014). However, without an operational definition of the theoretical construct the *IT-MAIS* is assessing, it is difficult to determine the assessment's validity beyond face value. The risk of using an assessment that is not valid or reliable could result in clinical providers making intervention decisions based on erroneous or missing information, thus limiting opportunities for intervention. We propose that additional psychometric analysis of the *IT-MAIS* could provide useful and needed information to assist clinicians and researchers in making assessment choices for very young children with SNHL.

Since the *IT-MAIS* was initially validated in 2000 (Zimmerman-Phillips *et al.*) researchers are using a new type of psychometric analysis—item response theory (IRT)—to develop new assessments and analyze existing ones (Engelhard 2013). In the following sections we provide our rationale for using IRT to analyze the *IT-MAIS* based on a sample of infant and toddler CI users and an overview of what the analysis entails.¹

¹We simplified the statistical explanation for the current readership. For a thorough discussion on the topic please see the tutorial written for those in the field of Communication Sciences and Disorders by Baylor and colleagues (2011).

Item Response Theory

Historically researchers used classical test theory (CTT) to establish the psychometric properties of assessments' validity and reliability. However, researchers in the field of education have used IRT modeling (Lord and Novick 1968) for the past four decades to develop assessments and analyze the psychometric properties of existing assessments (Engelhard 2013). Researchers are particularly motivated to use IRT because the resulting data provide information about the assessment at the item level. This is a benefit that cannot be obtained using CTT (Wright and Stone 1999) because it provides information at the level of the overall test in its entirety. CTT was the previous methodology used to validate the *IT-MAIS* (e.g., Zheng *et al.* 2009). For this study we echoed the choice of others in the field of audiology (Ng *et al.* 2016) and used Rasch analysis—one model based on IRT—to further explore the psychometric properties of the *IT-MAIS*.

Measuring a latent trait

IRT, also referred to as latent trait theory, is a paradigm used to design assessments with an emphasis on ensuring accurate test scoring and item development (An and Yung 2014). A *latent trait* is an underlying behavior that can be intuitively understood, but not directly observable (e.g., intelligence; Baker, 2001) and must be inferred based on a theoretically driven set of behaviors that represent the latent trait (Wright and Stone 1999). For the purposes of the present study, we named the latent trait measured by the *IT-MAIS listening development*. We operationally defined listening development as the hierarchical acquisition of sound detection, discrimination, identification, and comprehension, based on Erber's (1982) levels of listening.

IRT models can be one-dimensional or multidimensional. The model's delineation is based on the number of parameters a researcher wishes to investigate and model (Lord and Novick 1968); much of the research in the field of IRT employs one-dimensional models. For such one-dimensional models—referred to as a 1-parameter logistic (1-PL) IRT model—item difficulty is modeled based on person ability. At its most basic, one type of IRT model (Rasch analysis) converts ordinal data into interval data, and transforms person ability and item difficulty along a single interval scale. The benefits of this type of scaling are described below.

Rasch analysis

We chose Rasch analysis to assess the item-level psychometric properties of our longitudinal data derived from the *IT-MAIS* because we were most interested in understanding the difficulty hierarchy of the *IT-MAIS* items based on the sample's individual ability levels. Rasch analysis is also a logical choice for those working with low incidence populations (e.g., infants and toddlers with CIs) because it allows a researcher to study the item-level psychometric properties of an assessment using a smaller sample size (50–100 data points) than is required by the more complex IRT models (Linacre 1994). More recently Chen and colleagues (2013) reported that Rasch analysis of samples between 30–50 participants showed robust item fit which led them to suggest that Rasch analysis could be used in the case of rare diseases (i.e. low incidence populations such as pediatric CI users).

Rasch analysis also provides valuable information that cannot be gleaned from CTT, by transforming ordinal data into interval data rather than summing raw scores or reporting percent correct. Rasch is based on probabilistic modeling that a person of average ability will be able to perform an item of average difficulty 50% of the time. After much iteration the data are transformed into an interval scale along which person ability and item difficulty are calibrated. The result is an objective scale of a latent trait scale much as the ruler is an objective measure of length. The ruler is an interval scale where the units of measure are arranged smallest to largest. The unit of measure—the inch—is invariant; meaning the size of the unit is equal to every point along the ruler. Furthermore, units on an interval scale are additive. Therefore, for example, if one knows an object is 3" long, they know it is more than 1" long and less than 6" long. Objective measures are also sample free; meaning that one can measure the length of anything, not just specific items. Consider how assessments developed using the probabilistic modeling of Rasch analysis differs from assessments developed using CTT, where entire tests must be given to obtain scores, and scores on one test will not necessarily be obtained on a different test.

Finally, some researchers (Baylor *et al.* 2001) propose that Rasch analysis provides more clinical utility than CTT because the interval scale describes the latent trait ability as specific, observable behaviors, plotted along an item difficulty hierarchy. The item difficulty calibrated along the interval scale provides researchers and clinicians the ability to quantify change, compare a patient's performance on a given set of items at different time points (e.g., pre- and post-CI), or to compare one patient's performance to another at the same time point in a meaningful way. For the most part, summing raw scores or deriving percentages correct, as done in CTT, does not provide the same information (Engelhard 2013).

We propose that using Rasch analysis to assess the item-level psychometric properties of the *IT-MAIS* is a logical first step toward better understanding how well the *IT-MAIS* assesses listening development in young children who are deaf and use CIs. Rasch analysis is an IRT methodology that lends itself to investigating the item level psychometric properties of existing assessments with small sample sizes. Rasch analysis also produces an objective measure of a latent trait (i.e. listening development) which can provide useful information to researchers and clinicians who work with infants and toddlers pre- and post-CI. By answering the following experimental questions we will understand more about the validity, reliability and sensitivity of the *IT-MAIS*, as well as how it conforms to our operationally defined trait—listening development.

1. Does the *IT-MAIS* data meet the assumptions for Rasch analysis (i.e. unidimensionality and local independence)?
2. Does the *IT-MAIS* demonstrate item-level psychometric properties to adequately measure the latent trait: listening development?
3. Does the *IT-MAIS* separate the participating children into more than two levels of ability to adequately identify different functional levels of performance?
4. Does the Rasch-modeled *IT-MAIS* item difficulty hierarchy conform to the theoretical item difficulty hierarchy established *a priori*?

Materials and Methods

Participants

Parents of 23 CI users aged 10 to 36 months and receiving services from the University of Iowa Children's CI Program completed the *IT-MAIS* during regular visits to the Center for their children's CI candidacy assessments and post-CI care. All 23 children (12 male, 11 female) children were born to parents with normal hearing, and were identified with severe to profound bilateral SNHL within the first year of life. All parents reported spoken, American English as the primary language used at home. See Table 1 for the children's demographic data.

Assessment

Pediatric CI audiologists administered the *IT-MAIS* to each parent at least twice (once pre-CI and once post-CI), and most children were assessed additional times post-CI. Assessment interims ranged from 1 month to 1 year, according to scheduled administration at 2-month intervals during the first year after CI stimulation, then at 6-month intervals until 3 years post-CI. The actual intervals between assessment dates varied due primarily to missed visits and scheduling conflicts. See Figure 1 for each participant's schedule of repeated measures. Using these repeated measures, we analyzed a sample of 56 data points collected from the 23 parents.² The rationale for this analysis is discussed in the next section.

During *IT-MAIS* administration, the audiologist asked/explained each question and recorded the parents' responses on the designated response forms. The audiologist interpreted each parent's answer by using a Likert scale ranging from 0 (*never*) to 4 (*always*) for each question, regardless of the child's communication modality and the absence/presence of a listening device(s). If the question was unclear to the parent, the audiologist was permitted to recast the question using probes with alternative wording provided by the *IT-MAIS*. Finally, the audiologist who administered the *IT-MAIS* scored the *IT-MAIS* after its completion.

Psychometric Analyses

Exploratory factor analysis—Two assumptions—unidimensionality and local independence—must be met to perform Rasch analysis (Wright and Linacre 1989). To test the assumption of unidimensionality prior to conducting Rasch analysis, we conducted an exploratory factor analysis (EFA). Factor extraction was established at a minimum eigenvalue > 1.0, $\alpha = .05$. To meet the local independence assumption, we transformed the inter-item residual correlations into Fisher's z scores. In that form, we characterized local independence among items as 5% of the non-significant pairs with correlations ≥ 2 SD from the mean (Smith 2005).

Rasch analysis—We employed the Rasch polytomous rating scale using WINSTEPS 7.5 (Linacre 2010). The Rasch polytomous formula models the relationship between a participant's ability (i.e. trait level) and the probability of choosing each response category

²Rasch analysis is considered ideal for analysis for small samples ($N = 50$ – 100). We anchored the initial 23 scores according to Rasch methods and by so doing; the other 33 ratings are considered individual ratings during the probabilistic iterations used to estimate the model (Mallinson 2011). Our results are thus based on 56 person-level entries.

(i.e. 0 – 4 Likert scale of the *IT-MAIS*) for each item. The Rasch model for polytomous rating scales is represented by the following formula (Linacre, 1994):

$$\text{Log} (P_{nik}/P_{ni(k-1)}) * B_n - D_i - F_k$$

Where:

P_{nik} = the probability that person n , on encountering item I would respond (or be observed) in category k ,

$P_{ni(k-1)}$ = the probability that the response (or observation) would be in category $k-1$,

B_n = ability of person n ,

D_i = difficulty of item i

F_k = a rating scale threshold defined as the location corresponding to the equal probability of observing adjacent categories $k-1$ and k

Recall that Rasch analysis permits a researcher to examine individual items on an assessment at the level of item difficulty and person ability rather than at the total test score level.

The following paragraphs describe the various item-level psychometric information derived from Rasch analysis used to address our research questions and how they relate to traditional validity and reliability terminology used in CTT.

Item infit statistics: When developing an assessment it is critical that the questions/items are appropriate for the participants' ability levels. In other words, do the items capture or "fit" the behaviors of the latent trait they were developed to measure? Item fit statistics are used to determine whether individual test items fit the proposed Rasch model.

The ideal infit statistic would be 1.0, indicating that the Rasch-modeled responses and the actual responses on the assessment matched perfectly. For this study we chose to employ a common mean square infit statistic range of 1.4 to 0.6 with standardized z -scores > 2.0 ; a range frequently used in healthcare research (Wright and Linacre 1994). For our analysis, we removed misfitting items and rerun Rasch analysis until all items fell within the established infit criteria.

Item difficulty hierarchy: Since the *IT-MAIS* is used to document listening skill development in young children with SNHL, we also wanted to examine the *IT-MAIS*' item difficulty hierarchy. Ideally, if the item hierarchy represents a full range of ability (i.e. from most severely impaired listening ability to normal listening ability), floor or ceiling effects will be 10% and the items will typically range from -2 to $+2$ logits³. The logit is the interval scale unit of measure along the interval scale that results from Rasch analysis. It

³The logit is the interval unit of measurement used in Rasch analysis. It represents the relative differences between person ability and item difficulty that results during the log-odds transformation of the data based on the natural logarithm. The logit is calculated at ~ 2.718 (Wright and Stone 1999).

comes from a specific calculation, but can be thought of as the inch on the ruler example we presented in the introduction. Therefore, for a well-developed measure the listening development we would have items that are very easy for the most impaired respondents and very hard for the least impaired.

Item mean/person mean: Comparing the item M to the person M provides an indicator of internal consistency. When the 2 calibrated Ms have similar measures (i.e. 0 logits) it indicates that the items' difficulty hierarchy captured the range of person ability. In other words, we have items that capture everyone's ability levels. For this study an acceptable item M/person M match was set as the actual item M/person M (± 1 SD; Wright and Stone 1999). Note the item difficulty M will always be 0, which is the set point for Rasch modeling, defined as the probability of a person of average ability being able to complete an item of average difficulty 50% of the time (Wright and Stone).

Rating scale analysis: Because parents may reply to ordinal rating scales in unpredictable ways, depending on their understanding of the question, or the audiologist's interpretation of the *IT-MAIS* recommended "flexible interview format" (Zimmerman-Phillips *et al.* 2001), we employed rating scale analysis to determine whether parents used the *IT-MAIS* 0 – 4 Likert rating scale in a predictable way. Three criteria were established to assess the stability of the *IT-MAIS* rating scale system: 1) each rating category had to contain 10 observations; 2) the categories had to advance in a step-wise fashion from lowest to highest, and 3) outfit (i.e. outlier-sensitive fit) mean square < 2 . If the *IT-MAIS* 0 – 4 rating scale met the established criteria, it would demonstrate that the parents were using the 0 – 4 units in the way the developers intended (Linacre, 2002). If the scale failed to meet the established criteria, it would indicate that parents were not sensitive to some of the unit delineations, and the rating scale could be collapsed to better reflect how the parents used the units (Linacre).

Person reliability: The person reliability statistic is comparable to Cronbach's α , a measure of reliability reported in CTT. Cronbach's α reflects a measure of the relationship among test items. Thus, a high Cronbach's α would suggest that items have a close relationship and should be included in the same set. An acceptable Cronbach's α is in the range of 0.8 – 1.0. We established an acceptable person reliability statistic as 0.8.

Person separation: The person separation index is similar to the concept of sensitivity in CTT (i.e. can a test correctly identify the person with the problem). In Rasch analysis the person separation index represents an estimate of how reliably people responded to the questions based on their ability levels and indicates the number of different ability levels represented by the sample, in this case, different levels of listening development. A separation index > 2 would indicate the *IT-MAIS* reliably separated children into at least 3 statistically different ability levels. A separation index < 2 would indicate that the *IT-MAIS* items do not separate children into different levels of ability, thus it would not be particularly sensitive to ability levels.

Content validity: Finally, we established an a priori item hierarchy ranking to be compared to the final Rasch-modeled item hierarchy to determine the content validity of the *IT-MAIS*. This analysis is important because pediatric CI programs often use the *IT-MAIS* to measure

progress from pre- to post-CI, as if the assessment were organized in accordance with order of acquisition. However, the *IT-MAIS* authors do not report the assessment to be based on order of skill acquisition. Four graduate-level students studying communication sciences and disorders rank ordered the 10 *IT-MAIS* items based on their clinical experience and theoretical knowledge of listening development. All 4 students completed an undergraduate course in pediatric aural rehabilitation and had clinical experience with at least 1 pediatric CI user, but were not familiar with the *IT-MAIS* questions. Spearman's Rank Order correlation was used to determine the relationships among the raters' rankings using SPSS (IBM Corp. 2013).

Results

Question 1: Does the *IT-MAIS* data meet the assumptions for Rasch analysis?

First, we tested for adequate sample size using SPSS and the Kaiser-Meyer-Olkin value. The Kaiser-Meyer-Olkin value for the present data was 0.925 ("superb" according to Field (2009)), indicating that we had an adequate sample to complete the EFA. Bartlett's test of sphericity indicated that correlation between items was sufficiently large for EFA [$\chi^2(45) = 512.005, p < 0.001$]. Extraction was completed for eigenvalues > 1 with 25 iterations for convergence.

The scree plot in Figure 2 illustrates that only one factor (listening development) accounted for 71.36% of the variance for the 10 items. We concluded that the *IT-MAIS* items demonstrate unidimensionality and met the assumptions for Rasch analysis.

Table 2 presents correlation coefficients between each *IT-MAIS* item. Ideally, correlation coefficients should be within the range of 0.3 to 0.9. Based on these criteria, the correlation coefficients for the *IT-MAIS* items were sound. We tested local independence (i.e. no item responses are dependent on responses to other items) by transforming inter-item residuals (differences between observed and expected responses) to standardized units using Fisher's z -transformation procedure (Smith 2005). Fisher's z -transformed inter-item residual correlations indicated the items demonstrated local independence based on a range of z scores from -0.173 to $+0.120$ (Table 3), well within the established criteria $z \leq 2.0$. This test confirmed that our data met the second assumption for performing Rasch analysis.

Question 2: Does the *IT-MAIS* demonstrate item-level psychometric properties to adequately measure the latent trait (listening development)?

The following results provided the information needed to answer Question 2.

Item misfit—Based on misfit criteria (mean square within the range of 0.6 to 1.4; $z > 2.0$) the results showed that item 1 [*Is the child's vocal behavior affected while wearing his/her sensory aid (hearing aid or cochlear implant)?*] and item 10 [*Does the child spontaneously associate vocal tone (anger, excitement, anxiety) with its meaning based on hearing alone?*] exceeded these criteria. These results demonstrated that parents did not respond predictably to the 2 items. We eliminated the 2 misfitting items and completed all subsequent model estimates based on the 8 items that demonstrated acceptable infit criteria (see Table 4).

Person misfit—Person misfit is based on a series of iterations that Rasch analysis computes in accordance with parents' responses to other items around their children's ability levels. We adopted the same misfit criteria for persons that we used for items (1.4 mean square < 0.6; $z > 2.0$). Table 5 presents data from parents who did not predictably respond to the *IT-MAIS* items that were close to their children's predicted ability levels. During initial analysis of an instrument, misfitting items and persons may be retained or deleted depending on the researcher's needs (Wright 1999). Because this was a preliminary exploration of the *IT-MAIS*, we retained all data, as we had no way of knowing why these parents might have responded as they did. For example, parents may have responded unpredictably due to different audiologists using different examples to elicit responses. Parents may have responded unpredictably if they had a weak understanding of their child's listening behaviors and guessed on a question, but later answered that question differently and more in line with the child's ability. Lastly, it is possible that the *IT-MAIS* questions are not well worded or do not reflect observable behaviors parents could easily identify.

Visual representation of Rasch analysis (person-item map)—Results showed that many of the *IT-MAIS* items measured functional listening skills at the same level of difficulty as other items. They also showed that children's listening skills measured pre-CI were significantly lower than children's skills post-CI. In other words, the analyses showed that the item difficulty range was smaller than the person ability range (Figure 3), indicating that more difficult items are needed to assess the full range of the children's functional listening abilities.

Figure 3 shows a map of person ability and item difficulty where both variables are plotted on the same scale. *IT-MAIS* items 2 and 7 represent medium difficulty items because they were closest to 0 logit. Two items with the same logit measure may be considered redundant suggesting they measure the same level of the latent trait (i.e. listening development). The remaining *IT-MAIS* items measure the latent trait of listening development at different item difficulty levels. Rasch analysis dictates that if an assessment is psychometrically ideal (Wright and Stone 1999) the item difficulty should reflect a range of 3 to 4 logits (typically ranging from -2 to +2 logits). The item difficulty range for the *IT-MAIS* was ~1.5 logits, thus less than ideal.

Comparing person mean ($M = 0.8$ logits) to item mean ($M = 0$ logit) indicated that the match between item difficulty and person ability was adequate (person $M \sim$ item $M \pm 1$). Note, in Figure 3, person ability ranges from -6.0 to +3.6 logits—a wide range of ability represented by a relatively small sample. Thus, it can be concluded that the children we evaluated were a representative sample of the population the *IT-MAIS* purports to assess children with SNHL pre-and post-CI ages 0 to 3 years.

If 10% of the sample demonstrates either floor or ceiling effects, it is an indication that the items do not tap the full range of person ability levels. Based on the 10% criteria, the parents' reports did not demonstrate significant ceiling or floor effects (0% ceiling effect; 8.9% floor effect).

Person reliability—Person reliability represents the way in which parents with children of a given ability level respond to the test items. This means that a parent whose child demonstrated high-level listening skills reliably responded with the highest rating category (4/*always*) when responding to items of low-level listening skills. The criterion for acceptable person reliability index is 0.80 (comparable to Cronbach's α). In the present study, the person reliability index was 0.92, which is “highly acceptable” according to Rasch analysis.

Rating scale analysis—A sound rating scale must meet three requirements: 1) each rating scale category must contain at least 10 observations; 2) measures must advance linearly with each category; and 3) measures must have outfit mean square < 2 (Linacre 2002). The *IT-MAIS*' rating scale met all 3 of these criteria. The analysis demonstrated that parents did not consistently use all 5 categories of the *IT-MAIS*' 0 – 4 rating scale (see Table 6).

Question 3: Does the *IT-MAIS* separate the participating children into more than two levels of ability to adequately identify different functional levels of performance?

Person separation was 3.41 (> 2), demonstrating that the *IT-MAIS* separated person ability into at least 3 statistically different levels. That is, the analyses revealed distinct differences in listening abilities between children with profound SNHL pre-CI as compared with the children post-CI.

Question 4: Does the Rasch-modeled *IT-MAIS* item difficulty hierarchy conform to the theoretical item difficulty hierarchy established *a priori*?

There was a statistically significant, strong, positive correlation ($\alpha = .05$) between the graduate students' rankings and the modeled hierarchy [$\rho(6) = .903, p < .01$], which differs from the *IT-MAIS* item order. (See Table 7.)

Discussion

Globally, children are undergoing cochlear implantation at younger and younger ages (Colletti *et al.* 2012). This decline in age and the challenges associated with accurately assessing the functional hearing of infants raises concerns regarding the tools used to evaluate CI candidacy and post-CI progress. The present study focused on a popular (Uhler and Gifford 2014) parent-report tool developed with the intention to serve as a cohesive measurement of pre- and post-CI functional listening development—the *IT-MAIS* (Zimmerman-Phillips *et al.* 2001). The aim of this study was to further explore the psychometric properties of the *IT-MAIS* and provide researchers and clinicians with additional information when making decisions about using the *IT-MAIS* in their research and clinical practice. The implications of our study's findings are discussed below according to the analyses' results.

IT-MAIS data met the assumptions for Rasch analysis

We posed our first question to ensure that the data set met two critical assumptions important to studying a latent trait: that the items represented a unidimensional trait and that items

were independent of one another (local independence). Our results indicated that the *IT-MAIS* met these two critical assumptions, thus the items represented a single factor and they were locally independent.

***IT-MAIS'* item-level psychometric properties are not ideal for measuring the latent trait of listening development**

Content validity—We used Rasch analysis to analyze the item-level psychometric properties of the *IT-MAIS* because the analysis' results provide information similar to the traditional benchmarks of validity and reliability. Two of 10 *IT-MAIS* items (20%) were discarded from the final analysis because they did not meet the infit criteria (see Appendix for the list of items). As a result, content validity was brought into question. Misfitting items may indicate that a question is poorly worded or that it is not relevant to the listening development construct at all. For example, a parent's response to misfitting item 1 [*Is the child's vocal behavior affected while wearing his/her sensory aid...?*] is dependent on the young child's age at evaluation and their degree of hearing loss. Very young children are likely to vocalize in a manner similar to their normal-hearing peers up until ~ 9 months of age (Oller and Eilers 1988). Furthermore, a young child with residual hearing that affords them essential audibility of the speech signal is likely to vocalize in a manner similar to their peers with normal hearing, with or without their listening devices (Bass-Ringdahl 2010). This potential variability noted across the vocalizations of young children with SNHL is apt to contribute to the item's weak content validity and subsequent misfit.

Rasch analyses also revealed a number of parents who did not respond to questions reliably, based on person ability; specifically, 9 out of 56 data points exceeded misfit criteria. Excessive person misfit may indicate that the questions are not applicable to what the parents experience with their children, or that audiologists administered the *IT-MAIS* in ways that did not lead to consistent responses (e.g., used different probes, did not probe sufficiently).

Person ability and item difficulty—Our analysis of person ability—in comparison with item difficulty—also raised concerns regarding the validity of the *IT-MAIS*. Eleven out of 12 parents assigned their children, pre-CI, overall scores < 3.⁴ This marginal floor effect indicated that the items on the *IT-MAIS* did not assess these children's pre-CI listening abilities. Rather, these very low scores indicated that the children never displayed the behaviors addressed by the items (according to parent report). These results are not surprising given that the majority of our participating children were diagnosed at birth with severe-profound SNHL. The results have important implications for the future use of the *IT-MAIS* with young CI candidates who have greater residual hearing.

Prior to cochlear implantation, children are currently predicted to achieve scores of *0/never* on most of the *IT-MAIS* items because children with profound SNHL have very limited listening skills. However, there is a growing trend to implant children with more residual hearing, thus exceeding the U. S. Food and Drug Administration's current guidelines of

⁴Overall pre-CI *IT-MAIS* scores for our sample were as follows: 0 (n = 5), 1 (n = 2), 2 (n = 1), and 3 (n = 3), out of a possible overall score of 40.

bilateral pure-tone averages of 90 dB HL (e.g., moderate-severe SNHL). This research suggests that young children with residual hearing (and greater audibility of the speech signal) prior to implantation would score very differently on the *IT-MAIS* than the children in our current study (e.g., Gantz *et al.* 2000). The minimal range of sounds detectable to children with SNHL prior to CI receipt brings into question the use of the *IT-MAIS* as a measure for CI candidacy. Specifically, how many items can a child achieve a score greater than *0/never* and still be considered a candidate for CI surgery? Choosing CI surgery for a child is an important decision with irreversible effects that eliminate any residual hearing present before surgery and limit the child's chance to utilize future technology and/or medical advancements (e.g., hair cell regeneration). The current analysis of person ability in comparison with item difficulty suggests that the current version of the *IT-MAIS* does not demonstrate strong validity, and therefore, should not be used by itself to determine CI candidacy until the issues of validity are resolved.

Rating scale analysis—Our analysis indicated that the parents did not maximally use the *IT-MAIS* 0 – 4 Likert scale when rating their children's listening behaviors. Rating scales are employed to attain information about a participant's degree of skill rather than a basic *yes/no* or *right/wrong* distinction (Linacre 2002). If categories on a rating scale are not well defined and mutually exclusive, the reliability of the assessment is negatively affected (Linacre). The parents' irregular use of rating scale categories indicates that the categories are not properly calibrated in a step-wise manner (i.e. infrequent use of a score of 2 in the present study relative to the frequency of use for the other 4 scores (Linacre)). This finding is clinically relevant because parents' ratings are used to evaluate their children's listening skills (as opposed to a professional directly eliciting behavioral responses from a child).

One solution for improving the caregivers' use of the *IT-MAIS*' rating scale categories would be to alter the rating scale (e.g., reducing it to 4, instead of 5, categories) which would echo the parents' rating behaviors in the current study. However, because we did not obtain 10 responses per ranking category for each item altering the rating scale is ill advised at this point in time (Chen *et al.* 2013, Linacre 2002). A larger sample size ($N > 100$) would increase the likelihood of obtaining the needed number of observations per rating category ($n = 10$) to determine whether the rating scale met the established criteria.

***IT-MAIS* adequately identified different functional levels of listening development**

Reliability—Our results indicated that, the 8 *IT-MAIS* items used in the Rasch analysis were able to capture > 2 levels of person ability (3.41 levels) and represent a relative strength of the *IT-MAIS*. It is critical that a measure demonstrate sufficient person separation to track progress in skill development. Based on the 3.41 person separation index we propose that the *IT-MAIS* may be a viable starting point for the creation of a new assessment used to track listening development in children with SNHL. Specifically, researchers could utilize the participant separation demonstrated by the *IT-MAIS* as a guide for constructing new assessment items that address the full range of person ability.

***IT-MAIS* item order inconsistent with item order based on Rasch difficulty measures**

Further validation based on theoretical foundation—Recall that it is critical in the field of objective measurement that the construct one is measuring has a strong theoretical foundation. It is also important that the assessment's questions be designed to cover a full range of ability—from most basic behavior to most complex—to alleviate Type 1 and Type 2 errors. In the case of the *IT-MAIS*, item difficulty analyses indicated the order of *IT-MAIS* items was inconsistent with item order based on Rasch difficulty measures. These data suggested the current iteration of the *IT-MAIS* should not be viewed as a hierarchical progression of listening development. Thus it may not be the most ideal instrument upon which to determine functional listening development or establish optimal listening intervention in children with hearing loss. Until we have more definitive results, using another assessment, perhaps alongside the *IT-MAIS* might be the wisest course of action. For example, Bagatto, and Scollie (2013) suggested to initially use the *LittleEARS Auditory Questionnaire* (Coninx *et al.* 2003), an assessment designed to track listening development in CI users who were implanted by 24 months of age. Once the child reaches ceiling on the *LittleEARS* they would switch to the *Parents' Evaluation of Aural/Oral Performance of Children (PEACH)* (Ching and Hill 2005), a parent-report tool designed to assess listening and communication skills in children using hearing aids and/or CIs. Although neither of these parent-report tools has undergone rigorous psychometric analysis, unlike the *IT-MAIS* a bevy of research is emerging that supports their strengths across a variety of validity and reliability measures (Bagatto *et al.* 2011).

Alternatively, the item hierarchy we established *a priori* in accordance with Erber's work (1982) had a strong, positive relationship with the Rasch item order. This follow-up analysis suggested that if one were to reorder the current *IT-MAIS*' questions, to reflect a developmental listening hierarchy—like that proposed by Erber—it would likely strengthen and broaden the assessment's usefulness. Subsequently, making it possible not only to quantify listening development in CI users but also to account for individual differences across users and customize their device management and listening intervention. This finding adds a wrinkle to the discussion about using Rasch analysis for small sample sizes. Chen and colleagues (2013) reported that larger sample sizes (100 or 250) demonstrated more stable item parameters than smaller sample sizes (30 or 50). In fact, they reported that item parameters painted nearly opposite pictures of the item hierarchy. Our analysis appeared to confirm that the item hierarchy we established in this Rasch analysis was valid. The sample size controversy in Rasch analysis is not new and will continue. However, when dealing with low incidence populations, Rasch analysis may still provide valuable information, as it appears to have done within the present study.

Future directions

While our intent was to stimulate discussions about the *IT-MAIS*, we recognize that our results could be unsettling—particularly to pediatric CI professionals like those of Uhler and Gifford's (2014) aforementioned study and champions of the assessment. A change in implementation of care (or even sometimes the suggestion of change) is a challenge for any clinical practitioners (Cook and Odon 2013). However, it is our duty as researchers and

clinicians alike to adhere to the conscientious use of current best evidence in making decisions about patient care (Dollaghan 2007).

We can foresee possible future directions to better understand the psychometric properties of the *IT-MAIS* and subsequently improve the outcome measures that are available for young children with hearing loss. We suggest three possible paths. First, researchers could consider revising the *IT-MAIS* with two main goals: 1) develop new items and reword the existing items to assess an appropriate range of listening skills in pre- and post-CI users and 2) establish a new item difficulty hierarchy to reflect functional listening development.

Second, researchers could focus on exploring listening skills to establish a globally-accepted operational definition for listening development while conducting more theoretically-motivated research to move the field closer to a comprehensive model of listening and spoken language processing—for all types of listeners. Specifically, we propose including the role of cognitive and communication skills in the definition and understanding of listening development. This unification of cognition and listening is important given that listening is a complex, cognitive task that is still not fully understood (e.g., Jerger *et al.* 2013, Pichora-Fuller and Singh 2006). Furthermore, children with SNHL (and no additional disabilities) are likely to continue developing cognitively (pre-CI) prior to developing most listening skills. In contrast, children with normal hearing concurrently develop cognitive, language, and listening skills.

Third, we would like to re-analyze the *IT-MAIS* using a larger sample size ($n = 100$), in addition to analyzing the item-level psychometric properties of other assessments in pediatric CI programs' test batteries (e.g., the *LittleEARS Auditory Questionnaire* and the *PEACH*). Understanding more about these tools might allow us to develop an optimal comprehensive battery of assessments for tracking listening development pre- to post-CI.

Conclusions

In this study, we analyzed the item-level psychometric properties of the *IT-MAIS* via Rasch analysis to gain further understanding about its validity and reliability. We chose to analyze the psychometric properties of the *IT-MAIS* because very little information exists regarding its development and validation, although it is widely used to assess listening skills in children with SNHL ages 0 to 3 years pre- and post-CI. The results indicated that the *IT-MAIS* items demonstrated less than ideal psychometric properties and the *IT-MAIS* item order did not reflect the order in which children are expected to develop functional listening skills. Our findings suggest that there is a pressing need for further discussion among researchers and clinicians about 1) how the *IT-MAIS* is used, and 2) what other valid and reliable assessments could be used alongside or in place of the *IT-MAIS* to determine CI candidacy, establish treatment goals, or track progress in listening development in very young children with hearing loss.

Acknowledgments

The authors extend a big thank you to our colleagues on the University of Iowa's Cochlear Implant Program and all the families who volunteered their time for this study.

Biographies

Brittan A. Barker is an assistant professor with the division of Audiology in the Department of Communicative Disorders and Deaf Education, Emma Eccles Jones College of Education and Human Services at Utah State University. Dr. Barker has worked with children using cochlear implants for nearly 20 years and specializes in aural rehabilitation. Her multilayer research program is housed within her Spoken Language Processing lab where researchers work to better understand how people of all ages learn through listening. Specifically, Dr. Barker wants to understand how people can successfully learn about their world—even in situations when listening may be especially challenging (e.g., when listening to a teacher with a foreign accent or listening through a cochlear implant (CI) device). It is her goal to use the knowledge she gains from her research program to help establish listening programs for people with hearing loss and empower them to be effective communicators throughout their lives.

Neila J. Donovan is an associate professor and Director of the Communication Outcomes Research (COR) Lab in the Department of Communication Disorders at Louisiana State University. She sits on the American Heart Association Stroke Council, and serves as Chair of the Scientific Affairs Committee for the Academy of Neurologic Communication Disorders and Sciences. Dr. Donovan has been a practicing speech-language pathologist for more than 25 years, treating individuals with acquired neurogenic disorders of speech, language, and swallowing. Her research interests include developing treatments and treatment outcome measures for adults with acquired neurologic communication disorders, such as those with traumatic brain injury, stroke, dementia, and Parkinson's disease; and translating evidence-based treatments into treatments clinicians can use in their day-to-day practice.

Anne D. Schubert is a Speech-Language Pathologist on staff at Speech & Language Stimulation Center, Inc. in Fort Collins, CO. Ms. Schubert works to enable clients to communicate to their maximum potential in a respectful, enjoyable, and stimulating environment.

Elizabeth A. Walker is an assistant professor with the division of Audiology in the Department of Communication Disorders at the University of Iowa. Dr. Walker has worked with children using cochlear implants for nearly 15 years as a dual certified audiologist/speech-language pathologist and specializes in aural rehabilitation. Dr. Walker is the director of the Pediatric Audiology Lab; here she conducts a variety of research focusing on young children with hearing loss who use listening devices.

References

- An, X, Yung, Y-F. Item response theory: What it is and how you can use the IRT procedure to apply it. Paper SAS364-2014. Cary, NC: SAS Institute Inc; 2014.
- Bagatto MP, et al. A critical review of audiological outcome measures for infants and children. Trends in amplification. 2011; 15:23–33. [PubMed: 21873343]
- Bagatto MP, Scollie SD. Validation of the Parents' Evaluation of Aural/Oral Performance of Children (PEACH) rating scale. Journal of american academy of audiology. 2013; 24:121–125.

- Baker, FB. The basics of item response theory. 2nd. ERIC Clearinghouse on Assessment and Evaluation; 2001.
- Barker BA, Kenworthy MH, Walker EA. How we do it: Employment of listening-development criteria during assessment of infants who use cochlear implants. *Cochlear implants International*. 2011; 12:57–59.
- Bass-Ringdahl SM. The relationship of audibility and the development of canonical babbling in young children with hearing impairment. *Journal of deaf studies and deaf education*. 2010; 15(3):287–310. [PubMed: 20457674]
- Cardon G, Sharma A. Central auditory maturation and behavioral outcome in children with auditory neuropathy spectrum disorder who use cochlear implants. *International journal of audiology*. 2013; 52:577–586. [PubMed: 23819618]
- Chen WH, et al. Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of life research*. 2013; 23:485–493. [PubMed: 23912855]
- Ching, TYC, Hill, M. The parents' evaluation of aural/oral performance of children (PEACH) rating scale. Chatswood, New South Wales: Australian Hearing; 2005.
- Colletti L, Mandalà M, Colletti V. Cochlear implants in children younger than 6 months. *Otolaryngology--head & neck surgery*. 2012; 147:139–146. [PubMed: 22454156]
- Coninx, F, Weichbold, V, Tsiakpini, L. LittleEARS auditory questionnaire. Innsbruck: MED-EL; 2003.
- Cook BG, Odon SL. Evidence-based practices and implementation science in special education. *Exceptional children*. 2013; 79:135–144.
- Dollaghan, CA. The handbook for evidence-based practice in communication disorders. Baltimore, MD: Brooks Publishing Co; 2007.
- Engelhard, G. Invariant measurement: Using Rasch models in the social, behavioral, and health sciences. New York: Routledge; 2013.
- Erber, N. Auditory training. Washington DC: Alexander Graham Bell Association; 1982.
- Ertmer DJ, Jung J. Monitoring progress in vocal development in young cochlear implant recipients: Relationships between speech samples and scores from the Conditioned Assessment of Speech Production (CASP). *American journal of speech-language pathology*. 2012; 21:313–328. [PubMed: 22628109]
- Field, AP. Discovering statistics using SPSS. London: SAGE Publications; 2009.
- Gantz BJ, et al. Long-term results of cochlear implants in children with residual hearing. *Annals of otolaryngology, rhinology, and laryngology-supplement*. 2000; 185:33–36.
- IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp;
- Jerger S, et al. Effect of perceptual load on semantic access by speech in children. *Journal of speech, language, & hearing research*. 2013; 56:388–403.
- Kishon-Rabin L, et al. Developmental aspects of the it-mais in normal-hearing babies. *Israeli Journal of speech and hearing*. 2001; 23:12–22.
- Linacre JM. Sample size and item calibration stability. *Rasch measurement transactions*. 1994; 7:328.
- Linacre JM. Optimizing rating scale category effectiveness. *Journal of applied measurement*. 2002; 3:85–106. [PubMed: 11997586]
- Linacre, JM. Winsteps® (Version 7.5) [Computer Software]. Beaverton, Oregon: Winsteps.com.; 2010. Available from <http://www.winsteps.com/>
- Lord, FM, Novick, MR. Statistical theories of mental test scores. Reading, MA: Addison-Wesley; 1968.
- Mallinson T. Rasch analysis of repeated measures. *Rasch measurement transactions*. 2011; 25:1317.
- Ng IHY, et al. An application of Item Response Theory and the Rasch model in speech recognition test materials. *American journal of audiology*. 2016; 25:142–152. [PubMed: 27250898]
- Oller DK, Eilers RE. The role of audition in infant babbling. *Child development*. 1988; 59:441–449. [PubMed: 3359864]
- Pichora-Fuller MK, Singh G. Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation. *Trends in amplification*. 2006; 10:29–59. [PubMed: 16528429]

- Rasch, G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960/1980.
- Robbins AM, Renshaw JJ, Berry SW. Evaluating meaningful auditory integration in profoundly hearing-impaired children. *The American journal of otology*. 1991; 12:144–150.
- Smith EV. Effect of item redundancy on rasch item and person estimates. *Journal of applied measurement*. 2005; 6:147–163. [PubMed: 15795483]
- Uhler K, Gifford RH. Current trends in pediatric cochlear implant candidate selection and postoperative follow-up. *American journal of audiology*. 2014; 23:309–325. [PubMed: 25036461]
- Wright, BD. Fundamental measurement for psychology. In: Embretson, SE, Hershberger, SL, editors *The new rules of measurement: What every psychologist and educator should know*. Mahway, NJ: Erlbaum; 1999. 65–104.
- Wright BD, Linacre JM. Reasonable item mean-square fit value. *Rasch measurement transactions*. 1994; 8:370.
- Wright, BD, Stone, MH. *Measurement essentials*. 2nd. Wilmington: Wide Range, Inc; 1999.
- Zheng Y, et al. A normative study of early prelingual auditory development. *Audiology and neuro-otology*. 2009; 14:214–222. [PubMed: 19129696]
- Zimmerman-Phillips, S, Osberger, MJ, Robbins, AM. *Infant-toddler meaningful auditory integration scale*. Sylmar, CA: Advanced Bionics Corporation; 2001.
- Zimmerman-Phillips S, Robbins AM, Osberger MJ. Assessing cochlear implant benefit in very young children. *Annals of otology, rhinology, and laryngology -supplement*. 2000; 185:42–43.

Figure 2. Scree plot demonstrating no points of inflection; thus indicating there was only one factor (*listening development*).

Figure 3.

Map of Person ability and Item difficulty. Logit scale ranges from -6.0 to $+3.6$ for person ability and from -1.23 to $+0.52$ for item difficulty. Person ability mean is represented by the **M** to the left of the logit scale; item difficulty mean is represented by the **M** to the right of the logit scale (at 0 logits). Each **X** represents an individual child, **S** = 1 SD, **T** = 2 SD

Table 1

Demographic data for the 23 pediatric CI users.

ID	M/F	age at HL	age at CI	CI device type	strategy	CI ear	HA use	HL etiology	mother's ed
1	M	0	11	Nucleus CI 24 RE(CA)	ACE	B	none	Waardenburg Syndrome	coll
2	F	0	13	Nucleus CI 24 RE(CA)	ACE	B	B 11	CMV	post-grad
3	M	0	12	Nucleus CI 24 RE(CA)	ACE	B	B 6	hereditary (unspecified)	HS
4	M	0	16	Nucleus CI 24 RE(CA)	ACE	B	B 10	Auditory Neuropathy	coll
5	M	0	14	Nucleus CI 24 RE(CA)	ACE	B	B 7	unknown	post-grad
6	M	0	14	Nucleus CI 24 RE(CA)	ACE	B	B 11	unknown	post-grad
7	F	0	12	Nucleus CI 24 RE(CA)	HiRes	B	B 5	unknown	coll
8	F	0	22	Nucleus CI 24 RE(CA)	ACE	L	-	CMV	coll
9	F	0	13	Nucleus CI 24 RE(CA)	ACE	B	none	CMV	post-grad
10	M	0	16	Nucleus CI 24 RE(CA)	ACE	B	B 12	unknown	-
11	M	-	13	Nucleus CI 24 RE(CA)	ACE	B	-	-	-
12	F	0	13	Nucleus CI 24 RE(CA)	ACE	B	none	unknown	coll
13	M	-	12	Clarion HiRes 90K	HiRes	B	-	-	-
14	F	-	13	Nucleus CI 24 RE(CA)	ACE	B	-	-	HS+
15	F	0	9	Nucleus CI 512(CA)	ACE	B	B 6	Jervell Lange-Nielsen Syndrome	-
16	F	-	13	Nucleus CI 24 R(CA)	ACE	B	-	-	-
17	F	0	13	Nucleus CI 24 RE(CA)	ACE	B	B 5	unknown	HS+
18	M	2	13	Nucleus CI 24 RE(CA)	ACE	B	B 11	Connexin 26	-
19	M	-	14	Nucleus CI 512(CA)	ACE	B	B 7	-	-
20	F	-	13	Clarion HiRes 90K	HiRes	B	-	-	-
21	M	0	12	Nucleus CI 512(CA)	ACE	B	B 10	unknown	-
22	M	0	20	Nucleus CI 24 R(CA)	ACE	R	B 14	unknown	coll
23	F	0	17	Nucleus CI 24 R(CA)	ACE	B	B 11	-	coll

Note: **ID** = participant identification number; **M/F** = male or female; **age at HL** = participant's age when diagnosed with bilateral hearing loss (months); **age at CI** = participants' age at time of CI implantation (months); **CI device type** = cochlear implant device(s) used by the participant; **strategy** = CI processing strategy; **CI ear** = implanted ear (R = right CI, L = left CI, B = bilateral CIs); **CI use** = amount of time the participant used his/her CI relative to the time at testing (years; months); **HA use** = hearing aid use prior to implantation (months; R = right HA, L = left HA, B = bilateral HAs); **HL**

Table 2

Correlation coefficients between all 10 *IT-MAIS* items based on the factor model.

Item #	1	2	3	4	5	6	7	8	9	10
1	1.000	.554	.439	.471	.522	.671	.364	.513	.412	.516
2	.554	1.000	.679	.712	.711	.729	.701	.730	.637	.627
3	.439	.679	1.000	.882	.833	.719	.736	.717	.745	.667
4	.471	.712	.882	1.000	.824	.702	.740	.736	.675	.742
5	.522	.711	.833	.824	1.000	.807	.788	.793	.791	.674
6	.671	.729	.719	.702	.807	1.000	.690	.726	.719	.627
7	.364	.701	.736	.740	.788	.690	1.000	.718	.704	.679
8	.513	.730	.717	.736	.793	.726	.718	1.000	.743	.625
9	.412	.637	.745	.675	.791	.719	.704	.743	1.000	.612
10	.516	.627	.667	.742	.674	.627	.679	.625	.612	1.000

Table 3

Local independence of inter-item residual correlations for the 8 *IT-MAIS* items that demonstrated acceptable infit criteria.

<i>IT-MAIS</i> Item	Fisher's <i>z</i> transformation
2. Does the child produce well-formed syllables and syllable sequences that are recognized as "speech"?	0.026
3. Does the child spontaneously respond to his/her name in quiet with auditory cues only (no visual cues)?	−0.120
4. Does the child spontaneously respond to his/her name in the presence of background noise with auditory cues only (no visual cues)?	−0.092
5. Does the child spontaneously alert to environmental sounds in the home without being told or prompted to do so?	−0.059
6. Does the child spontaneously alert to environmental sounds in new environments?	0.121
7. Does the child RECOGNIZE auditory signals that are part of his/her everyday routines?	−0.173
8. Does the child demonstrate the ability to discriminate spontaneously between two speakers with auditory cues only (no visual cues)?	−0.034
9. Does the child spontaneously know the difference between speech and non-speech stimuli with listening alone?	−0.117

Table 4

Item infit statistics based on the established infit criteria for mean square (MnSq) and *z*-score for the 8 *IT-MAIS* items that demonstrated acceptable infit criteria.

Item	Measure	Model S.E.	Infit	
			Mnsq	<i>z</i>
2. Does the child produce well-formed syllables and syllable sequences that are recognized as “speech”?	0.49A	0.18	1.28	1.4
3. Does the child spontaneously respond to his/her name in quiet with auditory cues only (no visual cues)?	0.52A	0.18	1.29	1.5
4. Does the child spontaneously respond to his/her name in the presence of background noise with auditory cues only (no visual cues)?	0.10A	0.19	1.43	2.1
5. Does the child spontaneously alert to environmental sounds in the home without being told or prompted to do so?	−0.26A	0.19	1.09	0.5
6. Does the child spontaneously alert to environmental sounds in new environments?	−0.45A	0.20	1.07	0.4
7. Does the child RECOGNIZE auditory signals that are part of his/her everyday routines?	0.34A	0.18	1.16	0.9
8. Does the child demonstrate the ability to discriminate spontaneously between two speakers with auditory cues only (no visual cues)?	−0.13A	0.19	1.02	0.2
9. Does the child spontaneously know the difference between speech and non-speech stimuli with listening alone?	1.23A	0.24	0.98	0.0

Table 5

Listing of misfitting persons based on the established criteria for infit mean square (MnSq) and and infit z -score.

Participant ID #	infit MnSq	infit z
CI21	3.68	3.1
CI13	3.30	2.8
CI18	2.87	2.9
CI16	3.16	2.6
CI18	3.02	3.4
CI7	2.87	3.0
CI14	2.50	2.1
CI12	2.45	2.8
CI9	1.93	2.0

Table 6

Summary of category rating scale utilization criteria based on Category Rating Utilization Analysis for the 5-category rating scale for the 8 *IT-MAIS* items that demonstrated acceptable infit criteria.

Item	>10 Observations per Category	Average Measures Advance Linearly	Outfit MnSq <2
2. Does the child produce well-formed syllables and syllable sequences that are recognized as “speech”?	NO 0 = 13 *1 = 8 *2 = 8 3 = 16 4 = 11	YES -2.49 -1.15 0.81 2.17 2.36	NO *-3.26 1.72 *2.00 0.58 1.59
3. Does the child spontaneously respond to his/her name in quiet with auditory cues only (no visual cues)?	NO 0 = 14 *1 = 2 *2 = 5 3 = 14 4 = 21	YES -3.72 -0.64 0.21 1.64 2.21	NO *2.11 1.49 0.21 0.71 *2.64
4. Does the child spontaneously respond to his/her name in the presence of background noise with auditory cues only (no visual cues)?	NO 0 = 14 *1 = 8 *2 = 8 3 = 18 *4 = 8	YES -3.72 0.04 1.35 1.99 2.79	YES 1.89 1.28 0.60 1.58 0.39
5. Does the child spontaneously alert to environmental sounds in the home without being told or prompted to do so?	NO *0 = 8 *1 = 5 *2 = 7 3 = 18 4 = 18	YES -5.41 -3.90 -0.30 1.34 2.55	YES 0.39 0.75 0.76 0.62 0.46
6. Does the child spontaneously alert to environmental sounds in new environments?	NO 0 = 10 *1 = 5 2 = 11 3 = 23 *4 = 7	YES -3.10 -2.51 0.28 1.76 3.05	NO *4.23 0.98 0.97 1.33 0.27
7. Does the child RECOGNIZE auditory signals that are part of his/her everyday routines?	NO 0 = 17 *1 = 3 *2 = 6 3 = 17 4 = 13	YES -2.73 0.70 1.07 1.59 2.68	NO *2.74 1.38 0.46 *2.46 0.94
8. Does the child demonstrate the ability to discriminate spontaneously between two speakers with auditory cues only (no visual cues)?	NO 0 = 16 *1 = 6 *2 = 8 3 = 10 4 = 16	YES -2.89 -0.41 1.47 2.18 2.38	NO *2.65 0.49 0.33 1.10 1.25
9. Does the child spontaneously know the difference between speech and non-speech stimuli with listening alone?	NO 0 = 17 *1 = 4 *2 = 6 3 = 11 4 = 18	YES -2.63 0.04 1.08 1.78 2.40	NO *4.22 0.75 1.28 0.71 1.05

with misfitting persons removed; *indicates category rankings exceeding criteria for each item

Table 7

Item order based on *a priori* rankings from 4 MA-level speech-language pathology students for the 8 *IT-MAIS* items that demonstrated acceptable infit criteria.

Item
*5. Does the child spontaneously alert to environmental sounds in the home without being told or prompted to do so?
°6. Does the child spontaneously alert to environmental sounds in new environments?
°3. Does the child spontaneously respond to his/her name in quiet with auditory cues only (no visual cues)?
°9. Does the child spontaneously know the difference between speech and non-speech stimuli with listening alone?
°8. Does the child demonstrate the ability to discriminate spontaneously between two speakers with auditory cues only (no visual cues)?
°7. Does the child RECOGNIZE auditory signals that are part of his/her everyday routines?
+4. Does the child spontaneously respond to his/her name in the presence of background noise with auditory cues only (no visual cues)?
°2. Does the child produce well-formed syllables and syllable sequences that are recognized as “speech”?

Note:

- * item was ranked in the same position in both our *a priori* ranking and via Rasch item difficulty measures;
- ° item was ranked ± 1 rank position; and
- + item was ranked +3 rank positions in *a priori* hierarchy than in item difficulty order determined by Rasch analysis.