10-1-2019

Validity evidence for the littlEARS early speech production questionnaire: An english-speaking, canadian sample

Olivia Daub
The University of Western Ontario

Janis Oram Cardy
The University of Western Ontario

Andrew M. Johnson
The University of Western Ontario

Marlene P. Bagatto
The University of Western Ontario

Validity Evidence for the LittlEARS® Early Speech Production Questionnaire: An English-Speaking, Canadian Sample

Olivia Daub, Janis Oram Cardy, Andrew M. Johnson, Marlene P. Bagatto

The University of Western Ontario

Author Note

Olivia Daub, Graduate Program in Health and Rehabilitation Sciences, The University of Western Ontario; Janis Oram Cardy, School of Communication Sciences and Disorders, The University of Western Ontario; Andrew M. Johnson, School of Health Studies, The University of Western Ontario; Marlene P. Bagatto, School of Communication Sciences and Disorders and National Centre for Audiology, The University of Western Ontario.

Correspondence concerning this article should be addressed to: Olivia Daub, Graduate Program in Health and Rehabilitation Sciences, The University of Western Ontario, Elborn College, London, Ontario, Canada, N6G 1H1. Email: odaub@uwo.ca

Disclosures: This work was supported through funding provided by MED-EL to Dr. Marlene Bagatto (Principle Investigator) and Dr. Janis Oram Cardy (Co-Investigator).

Acknowledgements: We would like to thank Dr. Elizabeth Skarakis-Doyle for her thoughtful comments in developing the validity argument for the present work. We would also like to thank Jessica Tomarchio and Tessa Renken for their work in data collection and the Mary J. Wright Research and Education Centre at Merrymount for their assistance in recruiting participants.

Validity Evidence for the LittlEARS® Early Speech Production Questionnaire: An English-Speaking, Canadian Sample

Interest in the early spoken language productions of children who are deaf/hard-of-hearing (CDHH) has increased in recent years given the implementation of early hearing detection and intervention (EHDI) programs. Recognizing the importance of access to early linguistic information (whether spoken or signed) to future language development, EHDI programs follow benchmarks for the identification of CDHH within the first 3 months of life through universal newborn hearing screening and follow-up audiological assessment (Joint Committee on Infant Hearing, 2007; Moeller, Carr, Seaver, Stredler-Brown & Holzinger, 2013; Muse et al., 2013). With the identification of a hearing loss, EHDI programs support families in providing rich linguistic information to the CDHH in the modality (signed or spoken language) selected by the family. It has been documented that the majority of CDHH are born to families where one or both of the parents do not have a hearing loss and therefore communicate in the home using spoken language (Mitchell & Karchmer, 2004). The decisions regarding whether to pursue language in one modality over another, or both, are made by the family and supported by the program (Moeller et al., 2013). For families who elect to support spoken language development for their child, appropriate technological intervention (i.e., hearing aid amplification or, when appropriate, cochlear implantation) for CDHH has been shown to be a necessary component (Bagatto et al., 2016; McCreery et al., 2015; Tomblin et al., 2015). Spoken language supports (e.g., speech-language pathology services) and sign language supports (e.g., American Sign Language) are also routinely provided in EHDI programs. Improved spoken language outcomes for children who receive access to EHDI services have been demonstrated in several investigations (e.g., Ching & Dillon, 2013; Kennedy et al., 2006; Moeller, 2000; Tomblin et al., 2015; Yoshinaga-Itano, 2003). Recently, researchers have shifted away from evaluating EHDI efficacy overall, and toward identifying components of EHDI programs that are associated with differential improvements in outcomes, with the goal to refine EHDI services (Ching et al., 2013; Daub, Bagatto, Johnson, & Oram Cardy, 2017; Tomblin et al., 2015).

Monitoring progress in spoken language development during the earliest months of life is necessary to identify if or when CDHH begin to deviate from their peers with normal hearing. It has been documented that canonical babbling (production of well-formed syllables), the cornerstone of an infant's vocal development, is commonly delayed in CDHH and that prolonged delays are predictive of a continuing difficulty in learning spoken language (Moeller et al., 2007). However, monitoring spoken language development progress is immensely challenging, particularly in the clinical context. First and foremost, a lack of clinical tools designed to measure early spoken language development poses a significant barrier. Some tools have been designed specifically to explore early vocal development as a predictor for functional hearing (i.e., the Prelexical Infant Scale Evaluation; Kishon-Rabin, Taitelbaum-Swead & Segal, 2009) but evidence is lacking for whether the tool relates to other measures of spoken language development. Similarly, the Infant Monitor of Vocal Production (Cantle Moore & Colyvas, 2018), which was originally designed as a tool to educate parents on normal vocal development, has been documented to be sensitive to changes over time but is, at present, lacking evidence that the tool is related to other, accepted tools of spoken language. Finally, the Vocal Development Landmarks Interview (Ambrose, Thomas & Moeller, 2016) has been designed to measure preverbal vocal development to a child's first word combinations, using audio recordings of infants productions to support parent responding. The Vocal Development Landmarks Interview has been demonstrated to relate to the speech scale of the Mullen Scales of Early Learning

(Mullen, 1995), growth over time, and (for some scales) hearing loss characteristics (i.e., better-ear pure tone averages, Ambrose et al., 2016). However, all of these tools are in varying stages of development, and currently lack sufficient evidence to support clinical decision making.

The goal of the present study was to collect and appraise validity evidence for the English adaptation of the LittlEARS® Evaluation of Early Speech Production Questionnaire (LEESPQ) and its use as a spoken language progress monitoring tool from birth to 18 months of age. The LEESPQ was originally published in German and was designed as a companion tool to the LittlEARS® Auditory Questionnaire (Bagatto, Brown, Moodie & Scollie, 2011; Coninx et al., 2009) to support clinicians in an EHDI context to monitor spoken language. The LittlEARS® Auditory Questionnaire is currently used as a part of the University of Western Ontario Pediatric Amplification Monitoring Protocol (Bagatto, Moodie et al., 2011). In developing the monitoring protocol, Moodie and colleagues documented that pediatric audiologists found the LittlEARS® Auditory Questionnaire to be easy to score, administer, and interpret, and that it was preferred over other available tools (Moodie et al., 2011). The LEESPQ is similar in design to the LittlEARS® Auditory Questionnaire, with the aim to be clinically feasible: it is a short (27 yes-no questions) parent-completed questionnaire designed to capture major spoken language milestones in the first year and a half of life. To date, validity evidence has been collected for German and Turkish translations of the tool, and scale analyses of these versions suggest the tool appropriately captures the intended construct (i.e., spoken language in normally hearing children from 0-18 months; Keilmann, Friese, Lässig & Hoffman, 2018; Kosaner et al., 2014; Wachtlin, Brachmaier, Amann, Hoffmann, & Keilmann, 2017). The LEESPQ has also been demonstrated to be significantly correlated with age, and scores are unrelated to gender or multilingual status (Keilmann et al., 2018). To date, no such evidence has been collected for the English version.

The English LEESPQ was adapted from the German version with input from J.O.C., a registered speech-language pathologist, such that the spoken language milestones more closely align with English development.

Test validation is an iterative process wherein the necessary forms of validity evidence depend on the conclusions a clinician wishes to draw from a given test's results (*Standards of Psychological and Educational Testing*, American Educational Research Association, American Psychological Association, National Council on Measurement Education, 2014). In evaluating the psychometric appropriateness of adopting the LEESPQ for use in EHDI progress monitoring, it is important to first specify the intended score interpretation that the validity argument aims to assess (*Standards*, 2014). Foundational to any other score interpretations, the LEESPQ must meaningfully capture and represent spoken language production ability. If the LEESPQ does not accurately measure the spoken language construct, does so inconsistently, or does so incompletely, it cannot be used as an outcome measure of spoken language development. Therefore, as a part of the iterative validation process, the intended score interpretation evaluated in the present work was the following: LEESPQ® total scores and individual items represent meaningful quantification of spoken language development in young children from birth to 18 months of age. This validity argument, therefore, stipulates the following hypotheses:

1. The LEESPQ items consistently assess the intended construct, spoken language, demonstrating acceptable internal consistency and an appropriate factor structure (i.e., evidence based on internal structure).
2. The LEESPQ is associated with other spoken language assessments and, as a developmental measure, LEESPQ scores are significantly associated with age (i.e., evidence based on relations with other variables).

3. Items on the LEESPQ are sensitive to differences in spoken language development over time.

The present study evaluated this validity argument using a combination of scale and item analyses.

Method

Ethics and Informed Consent

This study was approved by Health Sciences Research Ethics Board (HSREB) at the University of Western Ontario. Parents provided written informed consent to participate in this study with their child, and parents served as reporters on behalf of their children.

Participants

A total of 103 children were tested to reach the final sample of 90 participants. Thirteen participants were excluded for the following reasons: otoacoustic emission refer result (n = 7), child too old to participate (n = 1), or concerns about the child's development (either through parent report or test results; n = 5). Participants in the final sample (N = 90) lived in the Southwestern Ontario, Canada area. Participants were recruited using a variety of strategies including word of mouth, advertisement through parent support and parent education programs, social media, community flyers, and a database containing contact information of parents who gave birth in local hospitals and agreed to be contacted about future research opportunities. Efforts were made to recruit a diverse sample, thus local agencies that support lower income families shared information about our study with families accessing their resources. Participants were equally sampled across nine age bins: 0-2 months, 3-4 months, 5-6 months, 7-8 months, 9-10 months, 11-12 months, 13-14 months, 15-16 months, and 17-18 months (see Table 1 for a summary). Seventeen children came from homes where languages in addition to English were spoken. Second language exposure varied from occasional songs or words in a second language, to up to 50% of the child's language exposure. Caregivers provided responses to the questionnaires about their child's development, 88 of whom reported to be the child's mother, one who reported to be the child's father, and one who reported to be the child's parent. Maternal education (in years, *M* = 17.01, *SD* = 2.84) was collected as a proxy variable for socio-economic status (SES) and linguistic environment, because it has been well documented that children from high SES homes tend to have higher scores on language assessments than children from lower SES homes, and that mothers with more education tend to provide richer linguistic environments (Ambrose, Walker, Unflat-Berry, Oleson & Moeller, 2015; Tomblin et al., 2015; Hart & Risley, 1995; Hoff, 2003). In the previously described 2 cases, the caregiver completing the demographic questionnaire provided their estimate of maternal education.

Materials.

Hearing screen. Children's hearing was screened by the first author or a trained research assistant using automated distortion-product otoacoustic emission (DPOAE) technology. Following an otoscopic examination, automated DPOAE screening was conducted, using the Madsen Accuscreen DP 5 protocol applied in Ontario's Infant Hearing Program (Herb & Derbyshire, 2013). Per this protocol, a refer result is indicated if the DPOAE signal to noise ratio is less than 8 dB on two or more frequencies. In Infant Hearing Program Screening, re-screening of an ear for which there was a refer result is permitted up to a maximum of two times. For the purposes of our study, children were not re-screened after a refer result, and re-screening was only initated if the DPOAE had to be terminated during the screen (e.g., if the child pulled the probe tip out of their ear). For this study, parent report of perceived hearing status was gathered to corroborate DPOAE screening.

Developmental screen. The *Nipissing District Developmental Screen* (NDDS) is a series of short screens used by physicians and other health care professionals in the province of Ontario to identify children at risk for developmental delays. The NDDS questionnaires cover major developmental milestones across different age ranges up to 6 years. The NDDS is a relatively broad developmental screen with low specificity (Dahinten & Ford, 2004), providing a conservative estimate to confirm the participants were meeting major developmental milestones.

***Demographic questionnaire*.** The demographic questionnaire was designed by the authors to collect basic information regarding the child's exposure to other languages, medical history, and maternal education.

Criterion language assessment. The *Receptive-Expressive Emergent Language Test, 3rd edition* (REEL-3; Bzoch, League & Brown, 2003) is a norm referenced, standardized parent interview measuring language development from birth to 3 years. The REEL-3 consists of two subtests (Expressive Language and Receptive Language) and questions are developmentally ordered. Normative data for the REEL-3 was collected on a sample of 1,112 children in the United States and demonstrates satisfactory validity evidence as a measure of spoken language development. The REEL-3 was selected as the criterion language assessment because of the availability of normative data, psychometric appropriateness, its measurement of both receptive and expressive language, and its suitability for the entire age range of children in our sample.

***LittlEARS® Early Speech Production Questionnaire*.** As previously described, the LEESPQ has 27 dichotomous (yes-no) items, and is a parent completed questionnaire. Items were designed to be developmentally ordered, beginning with items measuring behaviours present at, or shortly after, birth and concluding with items measuring behaviours that begin to emerge at 18 months in typically developing children. Items were based on English translations of the original German instrument. In developing the English forms, the test developers consulted with our team and J.O.C., an Ontario registered speech-language pathologist, provided feedback on the translation of the items prior to the beginning of our study.

Procedure. For each child who was recruited, hearing was screened in each ear using DPOAE technology and parent report of perceived hearing status was gathered to corroborate DPOAE screening results. In cases where the DPOAE could not be completed (e.g., the child was fussy and would not complete the testing), parent report that their child passed their newborn hearing screen and that they had no concerns regarding their child's hearing ability was accepted as confirmation of normal hearing status. In cases where children did not pass the DPOAE (n = 6), they were referred to the H.A. Leeper Speech and Hearing Clinic, the speech-language and audiology clinic at the University of Western Ontario, for additional assessment. The families were invited to complete the rest of the study after additional assessment confirmed normal hearing status. Two children were excluded and later re-included when audiological reassessment using visual reinforcement audiometry confirmed normal hearing status. The remaining four did not contact the examiner to provide updates, and it is unclear whether the families followed up with an audiological evaluation. One two-month-old child had a DPOAE result of refer, but the child's mother reported that the child had passed Automated Brainstem Response testing at birth. Given the proximity to an in-depth audiological assessment, the child's data were included in the analyses.

Following a pass on the hearing screening, caregivers completed the NDDS. In cases where a child was not meeting NDDS milestones, parents were informally probed by the examiner about milestones. Five children whose parents answered "No" to only one question on the NDDS were deemed to be eligible to continue with the study based on overly conservative responding on the part of the caregiver, and one child was included despite a “No” response because the caregiver had previously spoken to their pediatrician, who had no concerns. As an example of overly conservative responding, one parent reported that daycare providers (with whom the child spends most of the day) reported having observed the milestone, but the parent had not personally observed it and so they responded “No” to the question. In cases where there were concerns regarding whether the child was not achieving the milestone, parents were encouraged to speak to their pediatrician and the child’s data were excluded from the study. When communicative milestones were not met on the NDDS, testing proceeded given that the REEL-3 more specifically measured spoken language development. In these cases, J. O. C. followed up with the parent, and in the one case where there were continued concerns regarding spoken language development, the participant was excluded from analyses and parents were provided with contact information for government funded, preschool speech-language pathology services.

After completing the hearing and developmental screens, parents were interviewed using the REEL-3. Due to administration errors on this test involving incorrect calculation of basal and ceiling performance, two children were excluded from only those analyses involving the REEL-3. Finally, parents completed the LEESPQ as previously described.

Results

Total Score Validity Evidence: Internal Structure

To determine which analyses were most appropriate to evaluate the LEESPQ’s internal consistency, a factor analysis was undertaken to identify the underlying factor structure. Kaiser-Meyer-Olkin analyses were used to test factorability of the data (measure of sampling adequacy = 0.82). A measure of sampling adequacy that is 0.60 is considered *mediocre* evidence for factorability, and values 0.90 or higher are considered *marvelous* (Kaiser & Rice, 1974), and so this analysis supported further factoring of the data. To determine the number of factors to be extracted, parallel analysis was used, and this analysis suggested that there are two latent variables present in the current data set (see Figure 1). Common factor analysis with an oblimin rotation was used to explore how the LEESPQ items relate to the two latent variables identified by the parallel analysis. The items on the LEESPQ are designed to be developmentally ordered, and of those items that sufficiently load onto one of each factor, they can theoretically be conceptualized as items assessing *protophone development* and items measuring lexically meaningful vocalizations or *symbolic development*. Figure 2 displays the results from the factor analysis: items 1-4, as well as item 17, do not load on to either factor. Items 1 through 19 address both reflexive sounds and protophone development. Items 20 through 27 assess the use of vocalizations meaningfully or symbolically: imitating animal sounds, using phonetically consistent forms to refer to items (i.e., proto-words), and using full spoken words.

The dual factor structure of the LEESPQ suggests that Cronbach's alpha (α, one of the most commonly reported measures of internal consistency; Lance, Butts & Michels, 2006; McNeish, 2017; Tavakol & Dennick, 2011) might underestimate the internal consistency of the LEESPQ. In this case, Revell's total Omega (Ω, an estimate of internal consistency accounting for different factor structures; McNeish, 2017; Zinbarg, Revelle, Yovel, & Li, 2005) is a more appropriate estimate of the internal consistency of the LEESPQ. Acceptable internal consistency values are defined dependent on the consequences of the assessment results. Assessments used in basic psychological research that do not impact the care or management of the child might set a lower alpha standard (e.g., $\alpha = 0.60$). Manuals supporting the critical appraisal of psychometric tools have advocated for a minimum alpha value of 0.70 (Denman et al., 2017; Mokkink, Prinsen, Alonso, de Vet, & Terwee, 2018), however, others have advocated for higher standards ($\alpha = 0.90$ as minimally acceptable; Lance, Butts & Michels, 2006). As Omega is the more appropriate internal consistency measurement, $\Omega = 0.70$ is considered acceptable evidence that the LEESPQ items consistently relate to one another, but $\Omega = 0.90$ or higher is preferred. Internal consistency, as measured by Omega, for the LEESPQ is $\Omega = 0.92$, indicating strong evidence for the LEESPQ's internal consistency when its dual factor structure is taken into account.

Total Score Validity Evidence: Relations with Other Variables

Evidence from the factor analysis suggests that the items on the LEESPQ load onto two moderately correlated latent variables. Based on the content of the items, it is inferred that the latent variables are, theoretically, components of spoken language development (pre-linguistic and linguistic spoken language). These are, however, inferences based on data-driven analyses that require empirical evidence to confirm. Concurrent validity evidence, derived from examining the LEESPQ's relation with the REEL-3, could confirm whether the two latent variables pertain to the spoken language construct. High correlations between the raw scores of the LEESPQ and the raw scores of the REEL-3 Expressive Language subtest would be taken as strong evidence for concurrent validity. For the present validity argument, we considered correlations of 0.60 to be acceptable evidence for concurrent validity, with higher values indicating stronger evidence.

Pearson's product moment correlations associating the LEESPQ raw score and the REEL-3 expressive subtest raw score are highly, positively correlated (see Figure 3) indicating that increases in LEESPQ raw scores are associated with increases in REEL-3 expressive raw scores. These results suggest excellent concurrent validity ($r = 0.92$, $p < 0.01$), providing evidence that the LEESPQ captures the construct of spoken language. Similarly, Spearman's

Rho correlations between age (treated as a categorical variable, in months) and LEESPQ raw scores were highly positively correlated ($r = 0.90$, $p < 0.01$; see Figure 4), suggesting that increasing LEESPQ scores are associated with older ages with a wide range of normal variation. Sensitivity and specificity analyses, which would provide statistical evidence for the most appropriate standard deviation cutoff for the identification of scores that fall below age expectations, were beyond the scope of the present work. However, the Joint Committee on Infant Hearing considers CDHH to be making "appropriate progress" (Muse et al., 2013, p. e1334) when their language ability is "commensurate with or within 1 *SD* of their chronological age" (Muse et al., p. e1334), although other frequently used cutoffs are more conservative and range from 1.5 to 2 *SD*s below the mean (Spaulding, Plante & Farinella, 2006). Figure 4 also presents the range of variation when described using these *SD* cutoffs.

Individual Item Validity Evidence: Evidence from Item Response Theory

Traditional assessment best practice cautions speech-language pathologists against interpreting a child's performance on individual items (e.g., McCauley & Swisher, 1984). This tradition stems from the legacy of classical test theory, the psychometric paradigm that goverened much of early test development. One of the central tenents of classical test theory assumes that all items are equally predictive of the underlying construct being measured, and that no one question should be easier or harder than another. This statistical assumption is contrary to clinical intuition and clinical sense, but traditional validation work does not provide clinicians with statistical evidence to support their intuitions (Daub, Skarakis-Doyle, Bagatto, Johnson & Oram Cardy, 2019). Unlike the classical test theory, Item Response Theory (IRT) statistically evaluates how individual items on an assessment relate to the respondant's ability. IRT analyses, therefore, provide the statistical licensure to support clinical intuition (see Baylor et al., 2011 for an in-depth discussion).

Given the dual factor structure of the LEESPQ (see above), IRT models were evaluated separately for each factor (items relating to vocal development and items relating to symbolic development). Three models (a 1-parameter Rasch, a 2-parameter Rasch, and a latent-trait model) were evaluated for the items relating to both vocal development and symbolic development. 1-parameter Rasch models estimate the difficulty of each item, with discrimination constrained to be equal to 1. 2-parameter Rasch models estimate the difficulty of each item with an unconstrained discrimination (i.e., not necessarily equal to 1) that is set to be equal across all items. Latent traits models allow both difficulty and discrimination to vary across all items. For both factors, the latent trait model was the significantly better fit for the data using the Akaike Information Criteria (AIC) as an index of model fi (Vocal: AIC (16)= 1130.62, $p < 0.001$; Symbolic: $AIC(7)$=438.71, $p < 0.001$), therefore item difficulty and discrimination are not constant across all items on the LEESPQ.

As the latent trait model was the best fit for both factors, difficulty and discrimination parameters were derived from these models. Tables 2 and 3 summarize the difficulty and discrimination values of each item, and demonstrate that item difficulty broadly increases as the test progresses, suggesting that the items are reasonably developmentally ordered. However, this pattern is less clear for the Vocal development factor, and there are very large amounts of standard error for these items. This is unsurprising given the difficulty parents may have in reporting on very early vocal behaviours. These patterns are re-inforced when inspecting the item characteristic curves (see Figures 5, 6, and 7). Many of the items in the Vocal factor provide almost no information about a child's total score with the exception of items 5 (relating to happy vocalizations) and 16 (relating to canonical babble). Note that the information provided by Item 5 appears to be related to our inclusion of very young children (less than 1 month old) whose parents, if they said their child was not yet making happy sounds, also said their child was not yet producing more advanced sounds. In the symbolic factor (see Figure 7), the increasing pattern of difficulty is much clearer, and there is more information provided by the items in this factor than those in the Vocal factor.

Discussion

The data presented here provide evidence to support interpretations of the LEESPQ total scores as representing a valid quantification of a child's spoken language ability between the ages of 0-18 months. Factor analysis results indicate that the LEESPQ questions pertaining to protophone development and to meaningful word use load onto distinct latent variables that are only modestly correlated (see Figure 2), although evidence from relations with other variables (i.e., the REEL-3, $r = 0.92$) indicate that together these latent variables capture the spoken language construct. The LEESPQ also demonstrates high levels of internal consistency ($\Omega = 0.92$) when accounting for its factor structure, and is highly correlated with age ($r = 0.90$).

IRT data are presented to support additional interpretation of LEESPQ test results, above and beyond total score interpretations. Inspecting the difficulty values (Tables 2 and 3), it is clear that some items were exceptionally easy items and were endorsed by all parents (e.g., questions related to early oral motor development). Information about item parameters support clinicians in interpreting cases where a parent might respond "No" to questions where all other parents replied "Yes," indicating an abnormality in either responding or the child's development. Further, not all items on the LEESPQ provide information about a child's ability, especially items in the Vocal factor. This is unsurprising given the subtle advancements that are made in infant vocal development that do not correspond to adult phonological categories (Oller, 2000). A parent might not be able to detect or perceive their child's subtle acoustic advancements in protophone development, even though they are sensitive to more major changes, such as canonical babbling (Oller, Eilers & Basinger, 2001) and parent report is commonly used to measure lexical development in the first year and a half of life (Fenson et al., 1994). It might not be possible for a checklist of vocal behaviours to be able to provide nuanced information about an infant's prelinguistic vocal development, regardless of the quality of item writing.

Given the cross-sectional design of the present work, ability in our data can be loosely interpreted as developmental age because the children in our sample were typically developing. In pooling the responses cross-sectionally, we cannot dissociate differing levels of ability within an age group. Future work including larger samples of children at different ages, children with developmental disorders, or CDHH would be expected to allow dissociation of the contributions of age and ability to performance.

Limitations:. Some concerns with our data collection might limit the extent to which the evidence can be generalized to other populations. First, our sample size is too small to derive normative values and limits the generalizability of our findings to the broader population. In validity work examining the LEESPQ in other languages, sample sizes have exceeded 300 children (Keilmann, Friese, Lässig & Hoffman, 2018), with between 10 and 72 children in each age bin, where our study has 10 children per age bin. Similarly, the extent to which our findings generalize across the Canadian population is unclear. Second language exposure was not controlled, although previous work has demonstrated that total scores from the German version of the LEESPQ are unrelated to multilingual exposure (Keilmann, Friese, Lässig & Hoffman, 2018). Additionally, levels of maternal education in our sample is higher than the broader

Canadian population, with 71% of mothers who participated in our study having 16 years or more of formal education (the equivalent to a bachelor's degree or higher, although some parents in our study did complete multiple college degrees), as compared to 25% of all Canadian adults who have a university degree (Statistics Canada, 2011). Socio-economic status has been known to influence language development (Hart & Risley, 1995; Hoff, 2003) and parent responding on language assessments (Feldman et al., 2000; Fenson et al., 2000). Given that there is already a wide range of individual variation in total scores in our sample, it is likely that this variability will increase with increasing population representation. Whether the range of normal variation is, indeed, too wide to determine whether a child falls below age expectations is dependent on the cutoff selected and is prey to subjective selection (Spaulding, Plante & Farinella, 2006). Sensitivity and specificity analyses that determine the most appropriate cutoff score is beyond the scope of the present analyses and relies on having a gold standard criterion assessment that can identify children with language development difficulties at such young ages. To date, no such gold standard assessment exists for young children between 0-18 months.

Clinical Implications: Despite these limitations, our results provide useful for evidence for clinicians considering use of this tool. Results from the item analyses suggests that the LEESPQ is inappropriate to stage young children within protophone development until the onset of well-formed vocalizations (i.e., canonical babble), and is most informative when the child has begun to babble or use spoken language. Canonical babble is documented to be the first clear point of divergence in vocal behaviours between young CDHH and children with normal hearing (Ambrose et al., 2016; Moeller et al., 2007; Nathani Iyer & Oller, 2008; Oller, 2000) and prolonged delays have been demonstrated to be predictive of future language learning difficulties (Moeller et al., 2007). Although some work has posited that auditory based differences in vocalizations may exist in the earliest stages of protophone development (Kuhl & Meltzoff, 1996; Mampe, Friederici, Christophe & Wermke, 2009), there is currently no work to suggest that these differences are (a) present in young CDHH and (b) predictive of future language ability. As the items assessing canonical babble on the LEESPQ are highly informative to total test score, this is consistent with the research literature documenting typical vocal development. Therefore, clinicians using the LEESPQ must take care in avoiding interpreting typical vocal development as typical spoken language development overall.

The factor structure of the LEESPQ highlights the importance of accumulating additional validity evidence on a population of CDHH. The language learning difficulties experienced by CDHH are hypothesized to stem from their inconsistent access to auditory information (Moeller & Tomblin, 2015) rather than an impairment in the language learning mechanism broadly. These difficulties first observably manifest themselves in differences in canonical babble complexity and consonant inventories (Ambrose et al., 2016; Nathani Iyer & Oller, 2008), however, difficulties in vocal development continue to manifest themselves in lexical production *orthogonally* to their receptive abilities (Moeller et al., 2007). Because CDHH have difficulty in accessing auditory information, their vocal productions might be less complex, intelligible, or diverse and therefore might result in *No* responses for later items on the LEESPQ for different reasons than a young child with typical hearing or other developmental language concerns. In sum, the latent variable structure of a factor analysis of a population of CDHH might be quantitatively different than the factor analysis documented here. In such a case, clinicians must take care in using the theory of the underlying disorder to guide their interpretations of LEESPQ scores and differentiate whether a low score might be due to differences in symbolic representation or auditory access. Permanent childhood hearing loss does not rule out the possibility that the child will have additional language learning difficulties for reasons beyond auditory access. Theoretically teasing apart those difficulties expected to stem from inconsistent auditory access and those difficulties that stem from a broader language or developmental disorder is necessary for interpreting outcome data beyond whether a child is, or is not, performing within age-expectations. From a clinical perspective, it is important to be able to identify when children are, or are not, making the progress expected given their hearing loss to identify when changing the intervention plan is appropriate (Moeller et al., 2013, Muse et al., 2013). From an EHDI program-level perspective, these differentiations are important for the operationalization of program success and the setting of realistic expectations for spoken language outcomes. Whether or not the underlying cause for a difficulty with spoken language development (inconsistent auditory access or language learning difficulty more broadly) can (a) be dissociated and (b) dissociated using the LEESPQ could be evaluated by comparing and contrasting the item parameters and factor structure of the LEESPQ between children with normal hearing and CDHH. Differential item functioning analyses could be appropriate to identify which items, if any, on the LEESPQ (or another measure of spoken language development) are sensitive to teasing apart these difficulties.

Conclusion

The present work collected and appraised evidence designed to evaluate LEESPQ total score interpretations as representing a quantification of spoken language development in the first year and a half of life in an English learning, Canadian sample. The present data suggest that the LEESPQ holds promise as a clinically feasible and psychometrically appropriate tool. Future research should evaluate the evidence for additional score interpretations, including its sensitivity to differences between infants with typical and atypical development, in order to facilitate the LEESPQ’s adoption into clinical practice.

References

Ambrose, S. E., Thomas, A., & Moeller, M. P. (2016). Assessing vocal development in infants and toddlers who are hard of hearing: A parent-report tool. *Journal of Deaf Studies and Deaf Education*, *21*(3), 237–248. https://doi.org/10.1093/deafed/enw027

Ambrose, S. E., Walker, E. A., Unflat-Berry, L. M., Oleson, J. J., & Moeller, M. P. (2015). Quantity and quality of caregivers' linguistic input to 18-Month and 3-year-old children who are hard of hearing. *Ear and Hearing*, *36*, 48S–59S. https://doi.org/10.1097/AUD.0000000000000209

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.

Bagatto, M. P., Brown, C. L., Moodie, S. T., & Scollie, S. D. (2011). External validation of the LittlEARS Auditory Questionnaire with English-speaking families of Canadian children with normal hearing. *International Journal of Pediatric Otorhinolaryngology*, *75*(6), 815–817. https://doi.org/10.1016/j.ijporl.2011.03.014

Bagatto, M. P., Moodie, S. T., Brown, C., Malandrino, A. C., Richert, F., Clench, D., & Scollie, S. D. (2016). Prescribing and verifying hearing aids applying the American Academy of Audiology Pediatric Amplification Guideline: Protocols and outcomes from the Ontario Infant Hearing Program. *Journal of the American Academy of Audiology*, *27*, 188–203. https://doi.org/10.3766/jaaa.15051

Bagatto, M. P., Moodie, S. T., Malandrino, A. C., Richert, F. M., Clench, D., & Scollie, S. D. (2011). The University of Western Ontario Pediatric Audiological Monitoring Protocol (UWO PedAMP). *Trends in Amplification*, *15*(1), 57–76. https://doi.org/10.1177/1084713811420304

Baylor, C., Hula, W., Donovan, N. J., Doyle, P. J., Kendall, D., & Yorkston, K. (2011). An introduction to item response theory and rasch models for speech-language pathologists. *American Journal of Speech-Language Pathology*, *20*(3), 243–259. https://doi.org/10.1044/1058-0360(2011/10-0079)

Bzoch, K. R., League, R., Brown, V.L. (2003). *Receptive-Expressive Emergent Language Test*. Austin, TX: Pro-Ed.

Cantle Moore, R., & Colyvas, K. (2018). The infant monitor of vocal production (IMP) normative study: important foundations. *Deafness & Education International*, 1-17. https://doi.org/10.1080/14643154.2018.1483098

Ching, T. Y. C., & Dillon, H. (2013). Major findings of the LOCHI study on children at 3 years of age and implications for audiological management. *International Journal of Audiology, 52 Suppl 2*(November), S65-8. https://doi.org/10.3109/14992027.2013.866339

Ching, T. Y. C., Dillon, H., Marnane, V., …Yeh, A. (2013). Outcomes of early- and late-identified children at 3 years of age: Findings from a prospective population-based study. *Ear and Hearing, 34*, 535–552.

Coninx, F., Weichbold, V., Tsiakpini, L., Autrique, E., Bescond, G., Tamas, L., … Brachmaier, J. (2009). Validation of the LittlEARS® Auditory Questionnaire in children with normal hearing. *International Journal of Pediatric Otorhinolaryngology*, *73*(12), 1761–1768. https://doi.org/10.1016/j.ijporl.2009.09.036

Dahinten, V. S., & Ford, L. (2004). Validation of the Nipissing District Developmental Screen for use with infants and toddlers - Working Paper, *7437*. Retrieved from www.effectivepractice.org/site/ywd_effectivepractice/assets/pdf/2c_british_columbia_validation_study.pdf

Daub, O., Bagatto, M., Johnson, A., & Cardy, J. (2017) Language outcomes in children who are deaf/hard-of-hearing: The role of language ability before hearing aid intervention. *Journal of Speech, Language and Hearing Research, 60,* 3310-3320. https://doi:10.1044/2017_JSLHR-L-16-0222

Daub, O., Skarakis-Doyle, E., Bagatto, M.P., Johnson, A.M., & Oram Cardy, J. (2019). A comment on test validation: The importance of the clinical perspective. *American Journal of Speech-Language Pathology, 28*(1), 204-210. https://doi:10.1044/2018_AJSLP-18-0048

Denman, D., Speyer, R., Munro, N., Pearce, W.M., Chen, Y-W., & Cordier, R. (2017). Psychometric properties of language assessments for children aged 4-12 years: A systematic review. *Frontiers in Psychology, 8*, 1-28. https://https://doi.org/10.3389/fpsyg.2017.01515

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development*, *71*(2), 310–322. https://doi.org/10.1111/1467-8624.00146

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Measuring variability in early child language: Don’t shoot the messenger. *Child Development*, *71*(2), 323–328. https://doi.org/10.1111/1467-8624.00147

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethic, S. J., … Stiles, J. (1994). Variability in early communicative development. *Encyclopedia of Psychology, Vol. 7.*, *59*(5), 377–379. https://doi.org/10.1037/10522-163

Hart, B. & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes Publishing.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hyde, M., & Derbyshire, A. (2013). *Hearing screening protocol and support document.* Report prepared for the Ministry of Children, Community and Social Services Infant Hearing Program.

Joint Committee on Infant Hearing (2007). Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics, 120*(4), 898–921. https://doi.org/10.1542/peds.2007-2333

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, *34*, 111–117.

Keilmann, A., Friese, B., Lässig, A., Hoffmann, V. (2018). Validation of the second version of the LittlEARS Early Speech Production Questionnaire (LEESPQ) in German-speaking children with normal hearing. *European Archives of Oto-Rhino-Laryngology*, *275*(4), 883-888. https://doi.org/10.1007/s00405-018-4889-4

Kennedy, C., McCann, D., Campbell, M., Law, C., Mullee, M., Petrou, S., … Stevenson, J. (2006). Language ability after early detection of permanent childhood hearing impairment. *The New England Journal of Medicine*, *354*, 2131–2141. https://doi.org/10.1056/NEJMc061614

Kishon-Rabin, L., Taitlebaum-Swead, R., & Segal, O. (2009). Prelexical infant scale evaluation: From vocalization to audition in hearing-impaired infants. In L. S. Eisenberg (Ed.), *Clinical management of children with cochlear implants* (pp. 325–368). San Diego, CA: Plural Publishing.

Kosaner, J., Sonuguler, S., Ogut, F., Kirazli, T., Uruk, D., Ornekci, R., & Amann, E. (2014). LittlEARS early speech production questionnaire validation in hearing Turkish infants. *Cochlear Implants International*, *15 Suppl 1*, S39-42. https://doi.org/10.1179/1467010014Z.000000000164

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4 Pt 1), 2425–38. https://doi.org/10.1121/1.417951

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria. What did they really say? *Organizational Research Methods*, *9*(2), 202–220. https://doi.org/10.1177/1094428105284919

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology*, *19*(23), 1994–1997. http://https://doi.org/10.1016/j.cub.2009.09.064

McCauley, R.R., & Swisher, L. (1984). Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. Journal of Speech and Hearing Disorders, 49, 338 - 348

McCreery, R. W., Walker, E. A., Spratford, M., Oleson, J., Bentler, R., Holte, L., & Roush, P. (2015). Speech recognition and parent-ratings from auditory development questionnaires in children who are hard of hearing. *Ear and Hearing*, *36*(1), 60S-75S. https://doi.org/10.1097/AUD.0000000000000213

McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, (February). https://doi.org/10.1037/met0000144

Mitchell, R. E., & Karchmer, M. A. (2004). Chasing the mythical ten percent: parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies, 4*(2), 138-163. https://doi.org/10.1353/sls.2004.2005

Moeller, M. P. (2000). Early intervention and language development in children who are Deaf and hard of hearing. *Pediatrics*, *106*(3), 1–9. https://doi.org/10.1542/peds.106.3.e43

Moeller, M. P., Carr, G., Seaver, L., Stredler-Brown, A., & Holzinger, D. (2013). Best practices in family-centered early intervention for children who are deaf or hard of hearing: An international consensus statement. *Journal of Deaf Studies and Deaf Education*, *18*(4), 429–445. https://doi.org/10.1093/deafed/ent034

Moeller, M. P., Hoover, B. M., Putman, C. A., Arbataitis, K., Bohnenkamp, G., Peterson, B., et al. (2007). Vocalizations of infants with hearing loss compared to infants with normal hearing. Part II: Transition to words. *Ear and Hearing, 28*(5), 628–642.

Moeller, M. P., & Tomblin, J.B. (2015) An introduction to the outcomes of children with hearing loss study. *Ear and Hearing, 36,* 4s-13s. doi:10.1097/AUD.0000000000000210.

Moodie, S. T., Bagatto, M. P., Miller, L. T., Kothari, A., Seewald, R., & Scollie, S. D. (2011). An integrated knowledge translation experience: use of the network of pediatric audiologists of Canada to facilitate the development of the University of Western Ontario Pediatric Audiological Monitoring Protocol (UWO PedAMP v1.0). *Trends in Amplification*, *15*(1), 34–56. https://doi.org/10.1177/1084713811417634

Mullen, M. E. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service.

Muse, C., Harrison, J., Yoshinaga-Itano, C., Grimes, A., Brookhouser, P. E., Epstein, S., … Martin, B. (2013). Supplement to the JCIH 2007 position statement: principles and guidelines for early intervention after confirmation that a child is deaf or hard of hearing. *Pediatrics*, *131*(4), e1324–e1349. https://doi.org/10.1542/peds.2013-0008

Nathani Iyer, S., & Oller, K. D. (2008). Prelinguistic vocal development in infants with typical hearing and infants with severe-to-profound hearing loss. *Volta Review*, *108*(2), 115-138.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.

Oller, D. K., Eilers, R. E., & Basinger, D. (2001). Intuitive identification of infant vocal sounds by parents. *Developmental Science*, *4*, 49–60. https://doi.org/10.1111/1467-7687.00148

Spaulding, T. J., Plante, E., & Farinella, K. a. (2006). Eligibility criteria for language impairment: is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools*, *37*(1), 61–72. https://doi.org/10.1044/0161-1461(2006/007)

Statistics Canada. National Household Survey. (2011). *Education in Canada: Attainment, Field of Study and Location of Study.* (Catalogue no. 99-012-X2011001). Ottawa, ON: Statistics Canada.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Tomblin, J. B., Harrison, M., Ambrose, S. E., Walker, E. A., Oleson, J. J., & Moeller, M. P. (2015). Language outcomes in young children with mild to severe hearing loss. *Ear and Hearing, 36*, 76S–91S. https://doi.org/AUD.0000000000000219

Mokkink, L.B., Prinsen, C.AC., Alonso, J., de Vet, H.CW., & Terwee, C.B. (2018). COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMS). Manual prepared for Knowledge Center Measurement Instruments. Retrieved from https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf

Wachtlin, B., Brachmaier, J., Amann, E., Hoffmann, V., & Keilmann, A. (2017). Development and evaluation of the LittlEARS Early Speech Production Questionnaire - LEESPQ. *International Journal of Pediatric Otorhinolaryngology*, *94*, 23–29. https://doi.org/10.1016/j.ijporl.2017.01.007

Yoshinaga-Itano, C. (2003). From screening to early identification and intervention: Discovering predictors to successful outcomes for children with significant hearing loss. *Journal of Deaf Studies and Deaf Education*, *8*(1), 11–30. https://doi.org/10.1093/deafed/8.1.11

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ωH: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7

Tables

Table 1 Participant Characteristics

Age (months)		Maternal Education (years)		REEL-3 (Ability Score-Expressive Language)		REEL-3 (Ability Score-Receptive Language)	
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*
0 - 2	10	18.1 (12 – 23)	3.48	98.6 (89 – 112)	8.35	78.1 (67 – 100)	9.6
3 - 4	10	16.9(12 – 23)	3.07	101.5 (95 – 114)	5.89	87.3 (68 – 100)	12.57
5 - 6	10	16.5 (14-20)	3.31	99 (82 – 115)	11.04	91.3 (82 – 103)	8.07
7 - 8	10	17.4 (14 – 23)	2.41	103.22 (96 – 109)	4.65	96.78 (87 – 105)	5.4
9 - 10	10	15.65 (11-21)	2.97	99.8 (75 – 114)	11.27	103.1 (83 – 117)	11.78
11 - 12	10	17 (12 -22)	2.58	99.9 (92 – 112)	7.64	97.9 (77 – 107)	9.46
13 - 14	10	17.7(14 – 20)	1.83	101.2 (85 – 109)	7.84	99.6 (92 – 102)	6.09
15 - 16	10	16.6 (12 – 24)	3.34	107.6 (100 – 110)	9.58	104.7 (90 – 135)	13.55
17 - 18	10	17.25 (13.5 – 22)	2.66	104.33 (95-113)	6.44	104.37 (91 – 125)	12.97
All children	90	17.01 (11 – 24)	2.84	101.64 (75 – 115)	8.48	95.7 (67 – 135)	12.97

Table 2 Vocal Factor: Item Parameters

Item	Discrimination		Difficulty	
	Value	*SE*	Value	*SE*
1	0.203	1.147	-22.288	127.491
4	0.151	0.6687	-22.473	100.373
5	-37.115	3853.357	1.843	404.055
6	-1.599	0.540	1.599	0.799
7	-4.026	4.696	1.685	3.576
8	-2.488	0.749	0.821	0.482
9	-1.935	0.514	0.643	0.299
10	-2.584	0.870	1.217	0.684
11	-1.872	0.547	0.741	0.364
12	-4.105	1.496	0.397	0.239
13	-2.413	0.658	-0.344	0.212
14	-1.905	0.524	0.647	0.328
15	-1.909	0.487	0.030	0.171
16	-27.524	1679.578	0.618	79.095
17	-1.485	0.396	0.295	0.222
18	-3.935	1.243	-0.105	0.130
19	-3.669	1.470	0.252	0.177

Table 3: Spoken Language Factor: Item Parameters

Item	Discrimination		Difficulty	
	Value	*SE*	Value	*SE*
20	5.041	1.517	0.5725	0.363
21	3.214	0.936	0.6716	0.387
22	33.273	12706.012	0.6207	495.944
23	4.501	1.214	0.2160	0.150
24	5.696	3.181	0.7220	0.801
25	5.856	2.033	1.1395	0.753
26	18.870	360.629	1.4456	53.613
27	28.283	5776.795	1.3607	555.653

Figures

Figure 1: Parallel Analysis Scree Plot

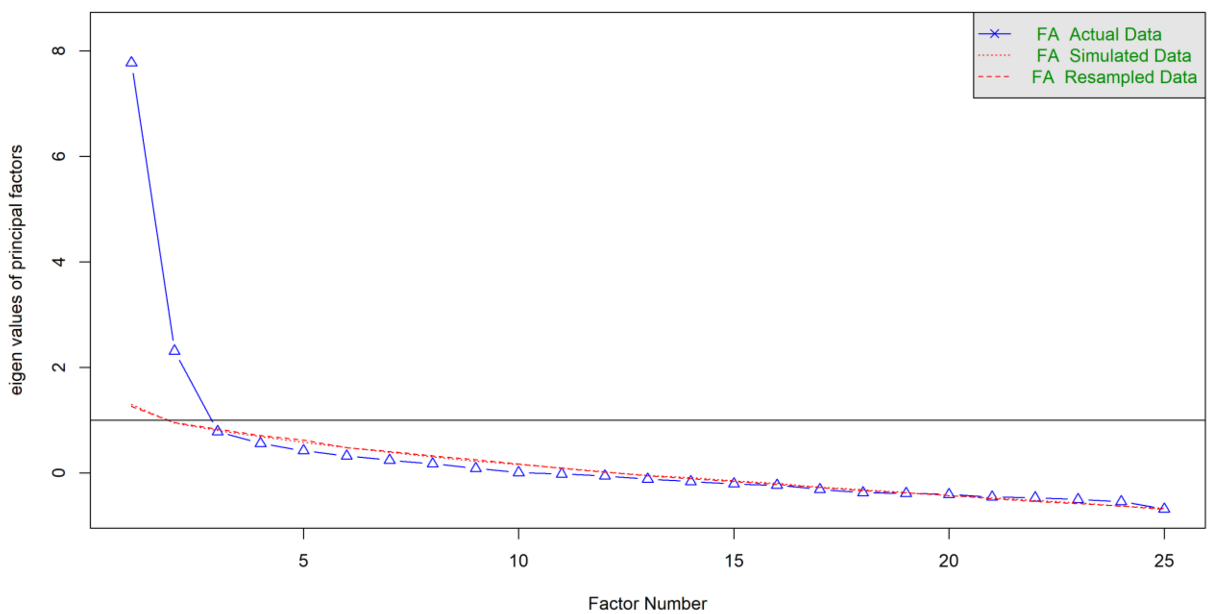

Figure 2: LEESPQ Factor Analysis

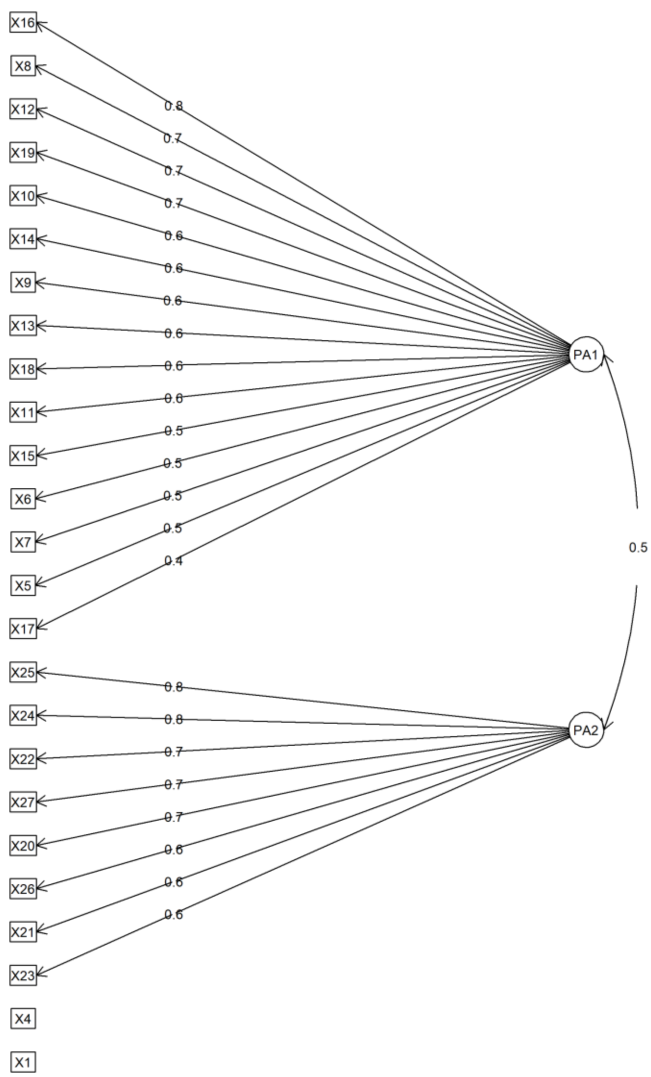

Figure 3: Relation between LEESPQ and REEL-3

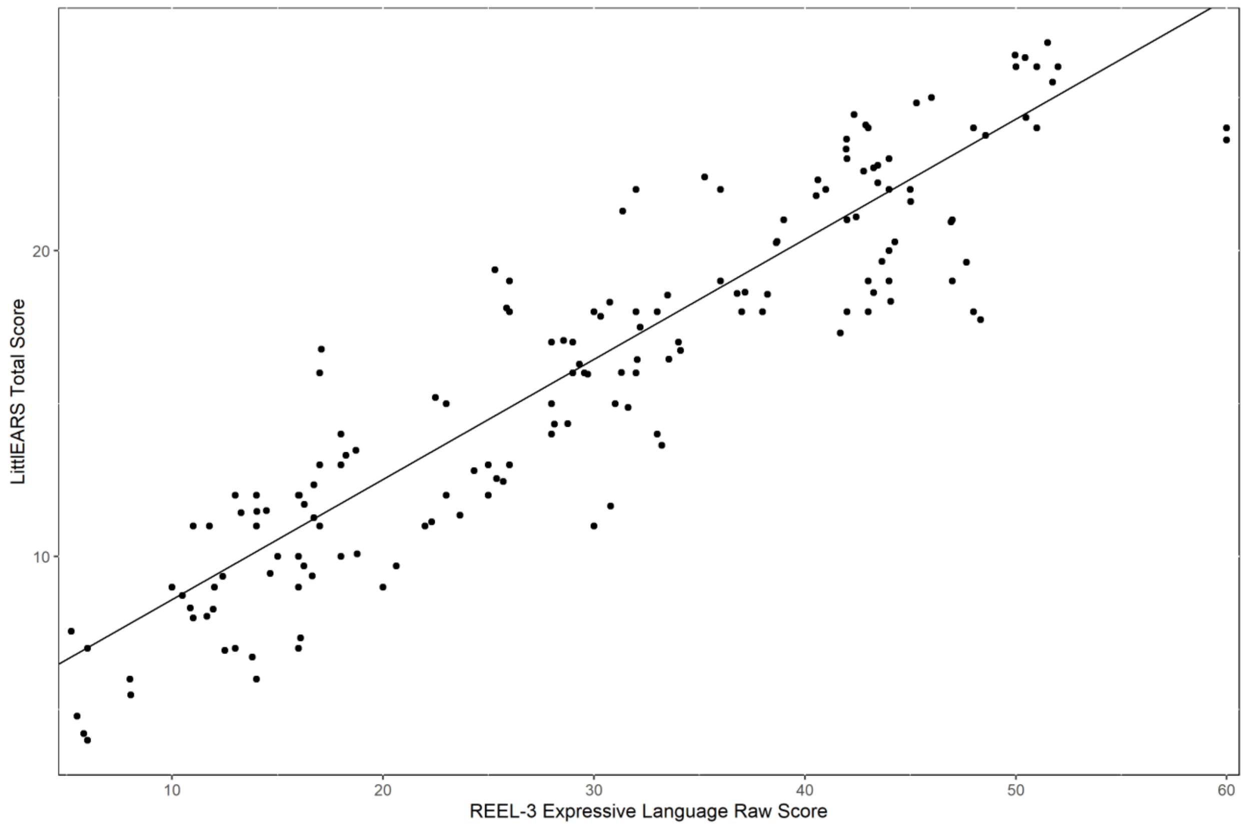

Figure 4: Relation between LEESPQ and Age

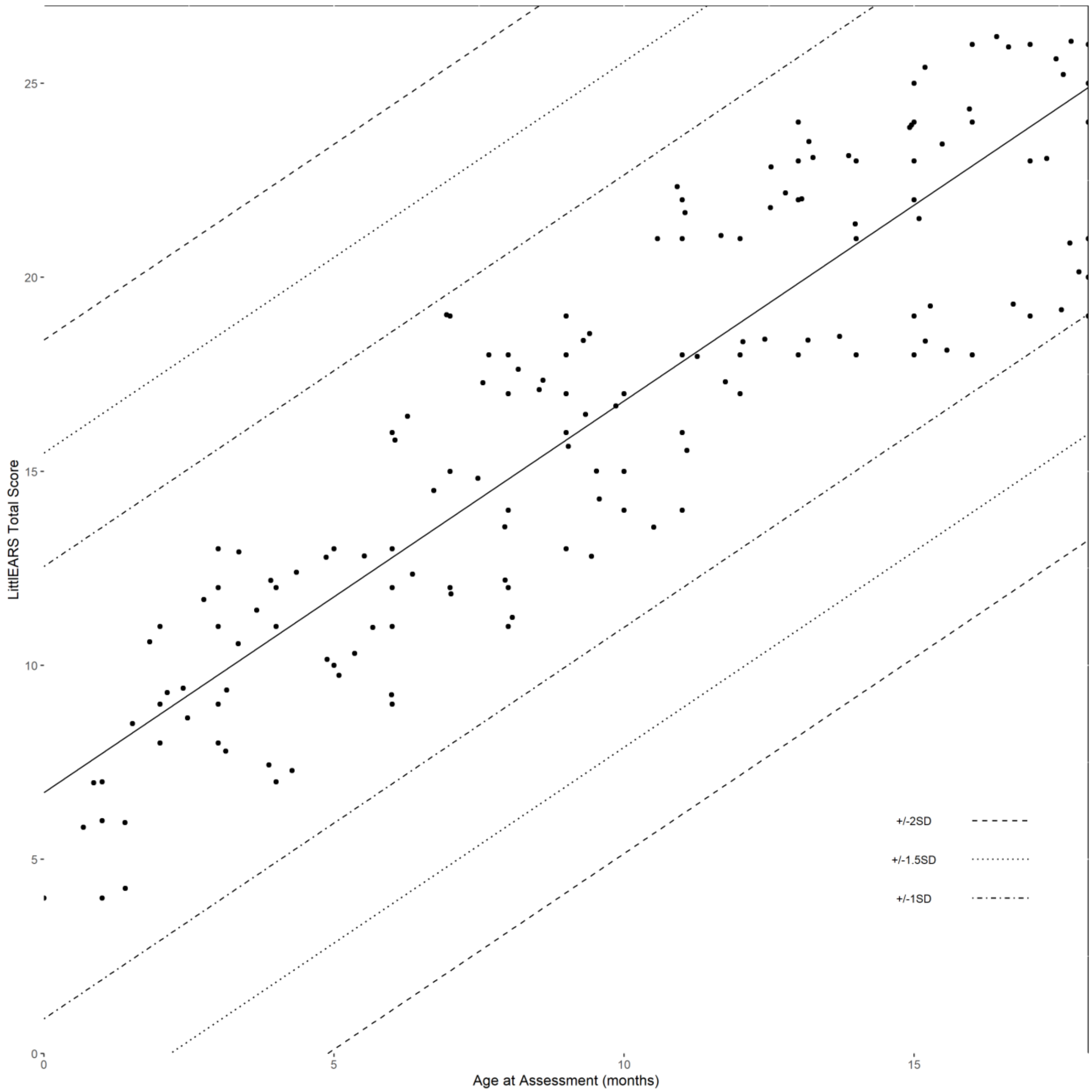

Figure 5: Vocal Development Information Curve (items 1–12)

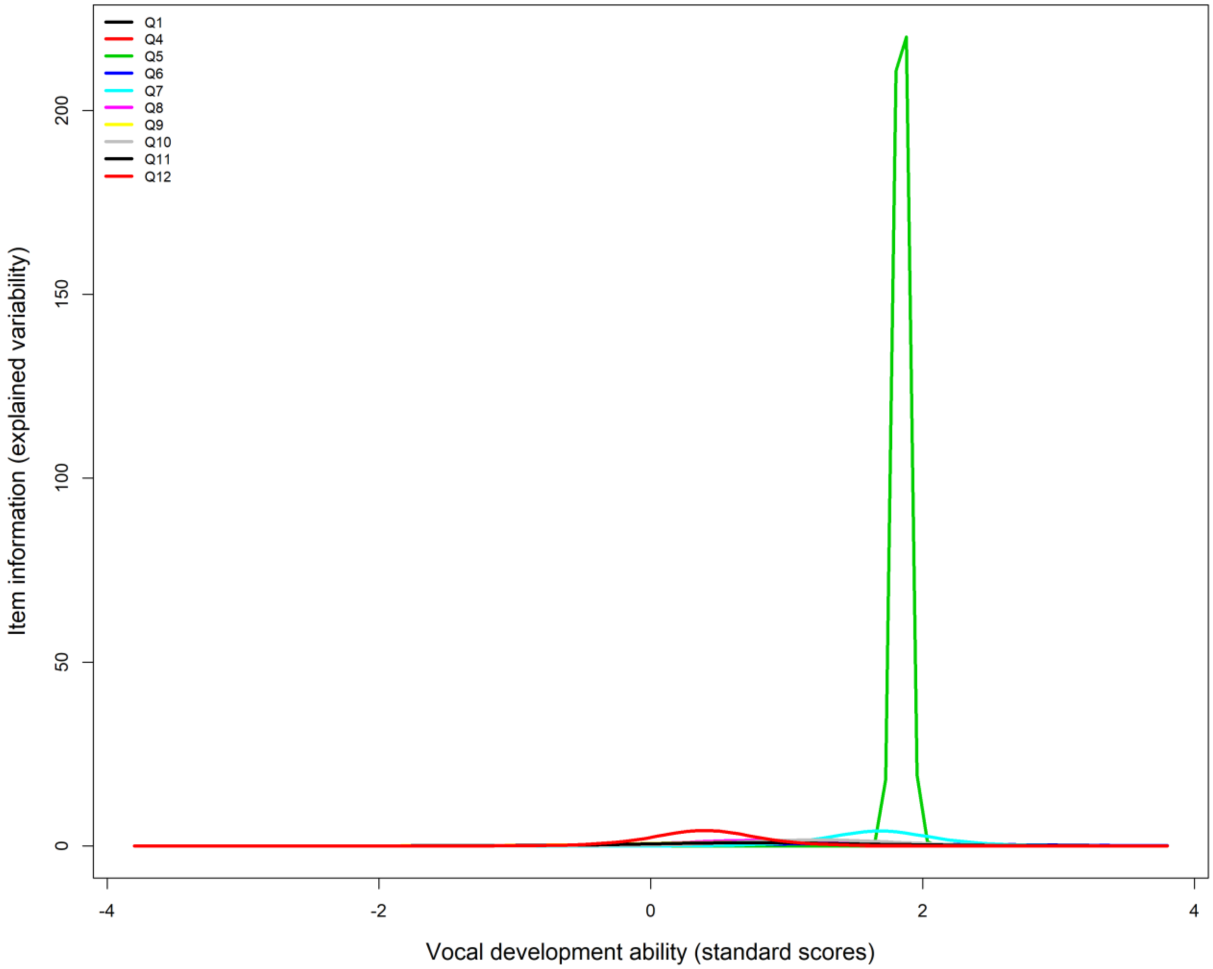

Figure 6: Vocal Development Information Curves (items 13-19)

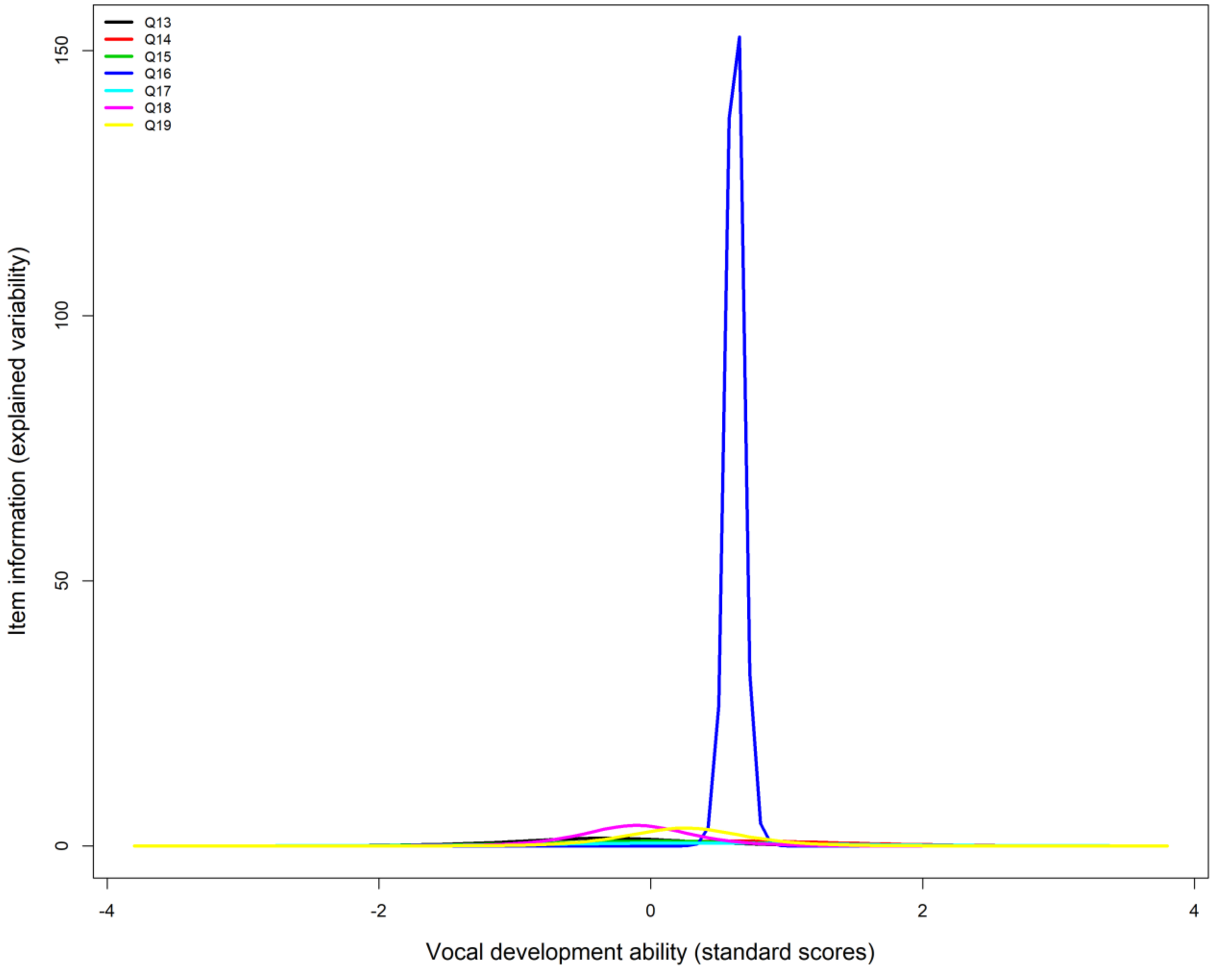

Figure 7: Spoken Language Development Information Curve