JEHDI
The Journal of Early Hearing Detection and Intervention
NCHAM National Center for Hearing Assessment and Management Utah State University™
2022; 7(1): 67–100

Usability and Feasibility of a Spoken Language Outcome Monitoring Procedure in a Canadian Early Hearing Detection and Intervention Program: Results of a 1-Year Pilot

Olivia Daub, PhD[1]
Barbara Jane Cunningham, PhD[2]
Marlene P. Bagatto, PhD[3]
Janis Oram Cardy, PhD[3]

[1]School of Epidemiology & Public Health, The University of Ottawa and Centre for Implementation Research, Ottawa Hospital Research Institute, Ontario, Canada
[2]School of Communication Sciences and Disorders, The University of Western Ontario, Canada
[3]School of Communication Sciences and Disorders and National Centre for Audiology, The University of Western Ontario, Canada

Abstract

Purpose: Best practice recommendations for Early Hearing Detection and Intervention (EHDI) programs include routine spoken language outcome monitoring. The present article reports on pilot data that evaluated the usability and feasibility of a spoken language outcome monitoring procedure developed for Ontario's Infant Hearing Program (IHP). This procedure included both Program-level monitoring using omnibus language tests from birth to 6 years of age and individual vulnerability monitoring of key domains of spoken language known to be at risk in children who are deaf or hard of hearing.

Methodology: Speech-language pathologists (SLPs) in the IHP piloted the new procedures for one year and provided feedback on the procedure through surveys at the end of the pilot.

Results: Data was suggestive that the Program-level procedure might be sensitive to change over time and known predictors of spoken language outcomes. Some, but not all, Program-level test scores were predicted by the presence of additional developmental factors. None of the test scores were significantly predicted by severity of hearing loss. Depending on the tests and scores used, some aspects of the Program-level procedure were sensitive to change over time. There was insufficient evidence to support individual vulnerability monitoring. SLPs reported significant concerns about the time involved in implementing both procedures.

Conclusions: This article describes preliminary evidence suggesting that the Program-level procedure might be feasible to implement and useful for evaluating EHDI programs. Future evaluations are needed to determine whether the procedure can be accurately implemented to scale in the IHP, and whether the data that results from the procedure can meaningfully inform stakeholders' decision-making.

Keywords: Spoken language outcome monitoring; Program evaluation

Acronyms: BEPTA = better-ear pure-tone average; CASL-2 = Comprehensive Assessment of Language Fundamentals, 2nd ed.; CELF-P2 = Clinical Evaluation of Language Fundamentals; dB HL = decibels Hearing Loss; DHH = deaf or hard of hearing; EHDI = Early Hearing Detection and Intervention; EOWPVT = Expressive One Word Picture Vocabulary Test; GFTA-3 = Goldman-Fristoe Test of Articulation, 3rd ed.; MBCDI-2 = MacArthur-Bates Communicative Development Inventories, 2nd ed.; IHP = Infant Hearing Program; IVT = Individual Vulnerability Test; Preschool – 2nd ed.; JCIH = Joint Committee on Infant Hearing; OMRU = Revised Ottawa Model of Research Use; PLS = Preschool Language Scale; PTA = pure-tone average; SII = Speech Intelligibility Index; SLP = speech-language pathologist

Correspondence concerning this article should be addressed to: Olivia Daub, PhD, School of Epidemiology & Public Health, The University of Ottawa and Centre for Implementation Research, Ottawa Hospital Research Institute, 501 Smyth Box 511, Ottawa, ON, K1H 8L6 Canada. Email: odaub@uwo.ca

Best practice recommendations for Early Hearing Detection and Intervention (EHDI) programs include routine spoken language outcome monitoring for infants who are born deaf or hard of hearing (DHH) and are learning a spoken language (Joint Committee on Infant Hearing [JCIH], 2007, 2013; Moeller et al., 2013). Routine spoken language outcome monitoring is intended to provide various stakeholders (i.e., administrators, clinicians, educators, families) with regular feedback on a child's development, and to support program evaluation and intervention planning. Stakeholders should expect that children who are DHH will progress toward age-appropriate spoken language outcomes regardless of the severity or type of hearing loss because hearing loss is not a language learning disorder (Moeller & Tomblin, 2015). Research has repeatedly demonstrated that when infants who are born DHH have adequate access to spoken language they perform, as a group, within age-expectations, but statistically below their

peers, on norm-referenced tests of overall spoken language ability (Ching et al., 2017; JCIH, 2019; Tomblin et al., 2015).

Despite the clear recommendations and rationale for spoken language outcome monitoring, there is limited evidence to support best practice recommendations for EHDI programs, and the clinical barriers and facilitators to implementing spoken language outcome monitoring procedures are not well understood. Daub and Oram Cardy (2021) provided the first report of the process used by one EHDI program, the Ontario Infant Hearing Program (IHP), to develop a standard spoken language outcome monitoring procedure. The IHP was launched in 2001 and provides comprehensive EHDI programming guided by JCIH best practice recommendations (JCIH, 2007, 2013, 2019). In the Canadian context, Ontario is one of the provinces/territories that continually provides adequate EHDI services through its IHP (Canadian Infant Task Force, 2014, 2019), including universal newborn hearing screening as well as intervention services to over 11,000 children who are DHH across the province annually. The IHP is a publicly funded program managed by the Ontario Ministry of Children, Community and Social Services. Children enrolled in the IHP access speech-language pathology supports from a related Ministry program—the Preschool Speech and Language Program, which is a publicly funded speech-language pathology program that serves all preschoolers in Ontario with speech, language, and communication needs. The Preschool Speech and Language Program employs more than 400 speech-language pathologists (SLPs) and provides services to more than 60,000 children with a wide range of needs (i.e., SLPs do not exclusively serve children who are DHH) each year. Preschool Speech and Language Program services are delivered in various contexts including designated clinics, childcare centers, in children's homes, and at fly-in clinics for families living in remote areas. The IHP previously tasked SLPs to use the Preschool Language Scale, 4th edition (Zimmerman et al., 2002) to monitor spoken language outcomes. Under this procedure, implementation between regions varied and SLPs tended to collect outcome data for children receiving IHP services *only* if the child was actively receiving Preschool Speech and Language Program services, that is, the outcome data were mostly focused on those children who were DHH for whom there were concerns about their spoken language development. When the Preschool Language Scale, 4th edition fell out of print, the IHP contracted the authors to support the development of a new procedure.

In developing a new spoken language outcome monitoring procedure, the authors and the IHP prioritized identifying a process for modelling growth in spoken language using norm-referenced tests that had previously been used in the peer-reviewed literature to evaluate children who are DHH. Based on the results of a scoping review, critical appraisal, and consultation with IHP managers and SLPs, a two-tiered assessment approach was recommended (Daub & Oram Cardy, 2021). In Tier 1, it was recommended that SLPs measure spoken language every six months from birth to 3 years, and annually thereafter (JCIH, 2007, 2013). Between birth and 1 year 6 months, SLPs were advised to use the MacArthur-Bates Communicative Development Inventories, 2nd edition (MBCDI-2; Fenson et al., 2007) Words and Gestures form, and from 1 year 7 months to 6 years, the Preschool Language Scale, 5th edition (PLS-5; Zimmerman et al., 2011) was recommended. The PLS-5 was selected based on its suitability for children within IHP age eligibility (up to 6 years), its psychometric appropriateness, and its Growth Scale Values, which are more sensitive to measuring change in language abilities than traditional norm-referenced scores (i.e., standard scores; Daub et al., 2017). Initial recommendations included using the PLS-5 right from birth, but concerns voiced by various stakeholders about the long administration time, lower diagnostic accuracy, and limited clinical value of the PLS-5 for children under 18 months of age, motivated the recommendation for use of the MBCDI-2 at the earliest ages. The purpose of the Tier 1 assessment was to collect data on children's spoken language outcomes that could be entered into a provincial database and used to facilitate program evaluation and planning (see Figures 1 & 2). Planned analyses for program evaluation included fitting growth curves of children's spoken language development and identifying factors predictive of growth in spoken language that could inform IHP curriculum development.

Figure 1

Overall Outcome Monitoring Process (from Daub & Oram Cardy, 2021)

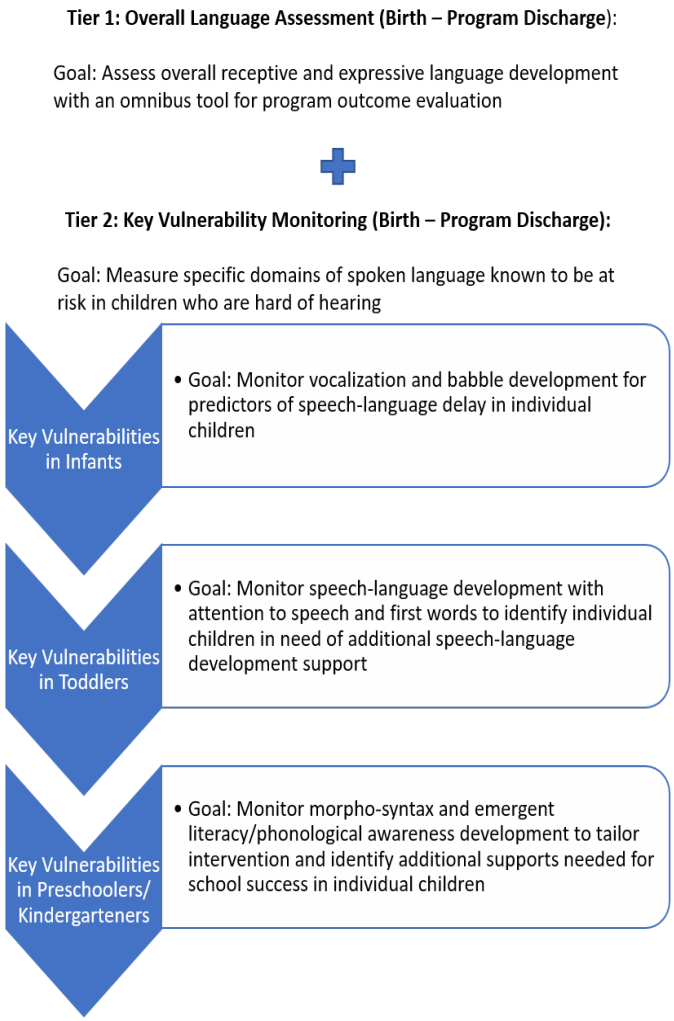


Figure 2

Tests Used in Outcome Monitoring Process (from Daub & Oram Cardy, 2021)

<table>
<tr><th></th><th>Program Monitoring</th><th colspan="3">Individual Vulnerability Testing</th></tr>
<tr><th>Age (years)</th><th></th><th>Vocalization/Babbling/ Articulation/Phonology</th><th>Words/Grammar</th><th>Emergent literacy/ Phonological awareness</th></tr>
<tr><td>0.5–1</td><td rowspan="2">MBCDI-2 Words & Gestures* (Scores for: Words Understood, Words Produced, Phrases Understood, and Gestures Produced)</td><td rowspan="3">Vocal development tests require further evaluation</td><td rowspan="2">(MBCDI-2 Words & Gestures)</td><td rowspan="4"></td></tr>
<tr><td>1–1.5</td></tr>
<tr><td>1.5–2</td><td rowspan="5">PLS-5 (Scores for: Auditory Comprehension & Expressive Communication)</td><td rowspan="2">MBCDI-2 Words & Sentences <u>or</u> EOWPVT-4</td></tr>
<tr><td>2–2.5</td></tr>
<tr><td>2.5–3</td><td rowspan="2">GFTA-3 (Scores for Sounds-in-Words)</td></tr>
<tr><td>3–4</td><td rowspan="3">CELF-P2 (Scores for Word Structure) <u>or</u> CASL-2 (Scores for Grammatical Morphemes)</td></tr>
<tr><td>4–5</td><td rowspan="2"></td><td rowspan="2">CELF-P2 (Scores for Pre-literacy Rating Scale) <u>or</u> CELF-P2 (Scores for Phonological Awareness Subtest)</td></tr>
<tr><td>5–6</td></tr>
</table>

Note. GFTA-3 = Goldman-Fristoe Test of Articulation, 3rd ed.; CASL-2 = Comprehensive Assessment of Spoken Language; CELF-P2 = Clinical Evaluation of Language Fundamentals, Preschool – 2nd ed.; MBCDI = MacArthur-Bates Communicative Development Inventories; PLS = Preschool Language Scale; EOWPVT = Expressive One Word Picture Vocabulary Test.

In Tier 2, it was recommended that SLPs assess key spoken language domains for which children who are DHH are at ongoing risk due to limitations with auditory access (see Figures 1 & 2). This tier was recommended as an improvement to the existing common practice whereby children were discharged from services when SLPs and families were not concerned about spoken language development. Tier 2 monitoring was recommended because permanent childhood hearing loss imposes lifelong limitations to auditory access, and it is therefore possible that delays in spoken language could still emerge despite overall age appropriate spoken language development being measured in a Tier 1 assessment. Tier 2 assessment recommendations included a list of tests SLPs could select from to measure each of three key individual vulnerabilities (see Figure 2). It was recommended that SLPs track key vulnerabilities at the same intervals as overall spoken language (every six months from birth to 3 years of age and annually afterward). For SLPs, the purpose of Tier 2 was to provide them with clinically useful information about a child's developmental status, facilitate intervention planning, and clarify the links between delays in different domains of spoken language development and overall spoken language performance. For the IHP, the purpose was to track key vulnerabilities to allow the program to model the development of three language domains for children who are DHH, and document agreement in disorder classification between omnibus spoken language assessments (Tier 1 MBCDI or PLS-5) and assessments specific to individual language domains (Tier 2 assessments).

Tier 1 and 2 recommendations were made based on the best available empirical and clinical evidence (Daub & Oram Cardy, 2021). However, evidence was still needed to confirm that these tiers resulted in usable data and were feasible to implement in the real-world. Although each of the tests included in the procedure were selected based on their alignment between psychometric properties and the IHP's program evaluation goals (Daub & Oram Cardy, 2021), it is possible that the data may not be sufficient when collected in practice. Whether the data can address the questions they are intended to answer depends on SLPs' ability to administer the procedure and enter the data into provincial databases. It must also be possible to extract the relevant data from the provincial database and prepare it for analysis.

This paper reports data from two pilot studies that were initiated to evaluate the usability and feasibility of both assessment tiers prior to program-wide implementation as a proof of concept that the data collected and entered conformed to our theoretical expectations when developing the procedure. These pilot projects were part of a series of program evaluation projects initiated by the IHP for which Western University provided methodological and statistical support. In Pilot Study 1, SLPs working in the IHP implemented the Tier 1 procedure for a one-year period and provided feedback through surveys on their perceptions of the procedure at the end of the pilot. In Pilot Study 2, a subset of SLPs from Pilot Study 1 simultaneously implemented the Tier 2 procedure and provided feedback at the end of the pilot. The current study addressed the following questions for the Tier 1 pilot:

1) Is the procedure sensitive to known predictors of spoken language outcome?
2) Is the procedure sensitive to change over time?
3) What are the barriers that SLPs experienced in implementing the procedure?
4) What modifications can be made to the procedure to improve its clinical feasibility?

The Tier 2 testing procedure was developed with the intention to provide information about key domains of spoken language to inform service provision for individual children. This study addressed the following questions for the Tier 2 pilot:

1) Does the procedure provide unique information beyond the Tier 1 procedure?
2) Do SLPs believe that the procedure is clinically useful?
3) What barriers did SLPs experience in implementing the procedure?
4) What modifications would improve the procedure's clinical feasibility?

Pilot Study 1: Tier 1 Program-level Outcome Monitoring

Method

Ethical Approval

Both pilots were Program Evaluation and Quality Improvement projects with the Ontario Ministry of Children, Community and Social Services. These projects were reviewed by the Western University Research Ethics Board (REB). The REB considered the projects not to be research as described in the Canadian Tri-Council Policy Statement V.2 (Research Exempt from REB Review, Article 2.4) and therefore they were not considered to fall under the purview of the REB.

Procedure

Prior to implementing the pilot program, participating SLPs ($N = 56$) from 11 regions in Ontario completed an online learning module designed to introduce and support implementation of the new spoken language outcome monitoring procedures (see Cunningham et al., 2021). SLPs implemented the recommended procedures in practice, routinely assessing the spoken language of all IHP children on their caseloads for one year (data collection completed in July 2019). At each assessment point, SLPs entered de-identified data into a secure REDCap database on a local server including test scores, age, and unique IHP identification number. SLPs also reported additional factors they believed influenced the child's scores (e.g., a comorbid diagnosis) or performance (e.g., distractibility). The first author (O.D.) then used the data in the REDCap database for analysis and checked all test scores for typographical or scoring errors by comparing the test scores SLPs entered into the database against the child's age using the examiner's manuals. Unique identification numbers were used to extract additional clinical information (i.e., child's sex, audiological variables) from the IHP database. The first author (O.D.) then used each child's identification number to link the demographic and audiological data with the pilot data. The final dataset was used to assess whether the procedures were sensitive to change over time and to predictors of spoken language outcomes. Note that the IHP database was managed by the IHP for clinical, not research, purposes and we did not have access to complete clinical charts or all variables that may impact children's language. Similarly, we did not have access to SLPs' clinical charts, and so we were unable to identify whether the data entered in the REDCap database represented *all* children on SLPs' caseloads who were eligible to be assessed with the procedure. These data, therefore, represented an opportunity to broadly investigate whether the outcome monitoring procedure conformed to our expectations.

To identify barriers to implementation and modifications required to improve feasibility, SLPs completed surveys designed to evaluate potential barriers to future implementation of the procedures at the end of the one-year pilot. Surveys were designed based on the Revised Ottawa Model of Research Use (OMRU; Graham &

Logan, 2004) and modelled after surveys used in the design of procedures to monitor auditory based outcomes for pediatric audiologists (Moodie et al., 2011). The OMRU is a framework to guide implementation of new innovations (in our case, spoken language outcome monitoring procedures) including assessing influential barriers and supports (i.e., features of the innovation, potential adopters, and the practice environment) related to implementing the innovation. Once implementation has begun, the OMRU recommends ongoing monitoring to generate evidence of the innovation's adoption and impact. Our feasibility analysis is positioned within the *assess* stage of the OMRU and our surveys were designed to understand factors about the innovation, potential adopters, and practice environment that may influence future implementation efforts.

Participants Assessed in Pilot Study 1

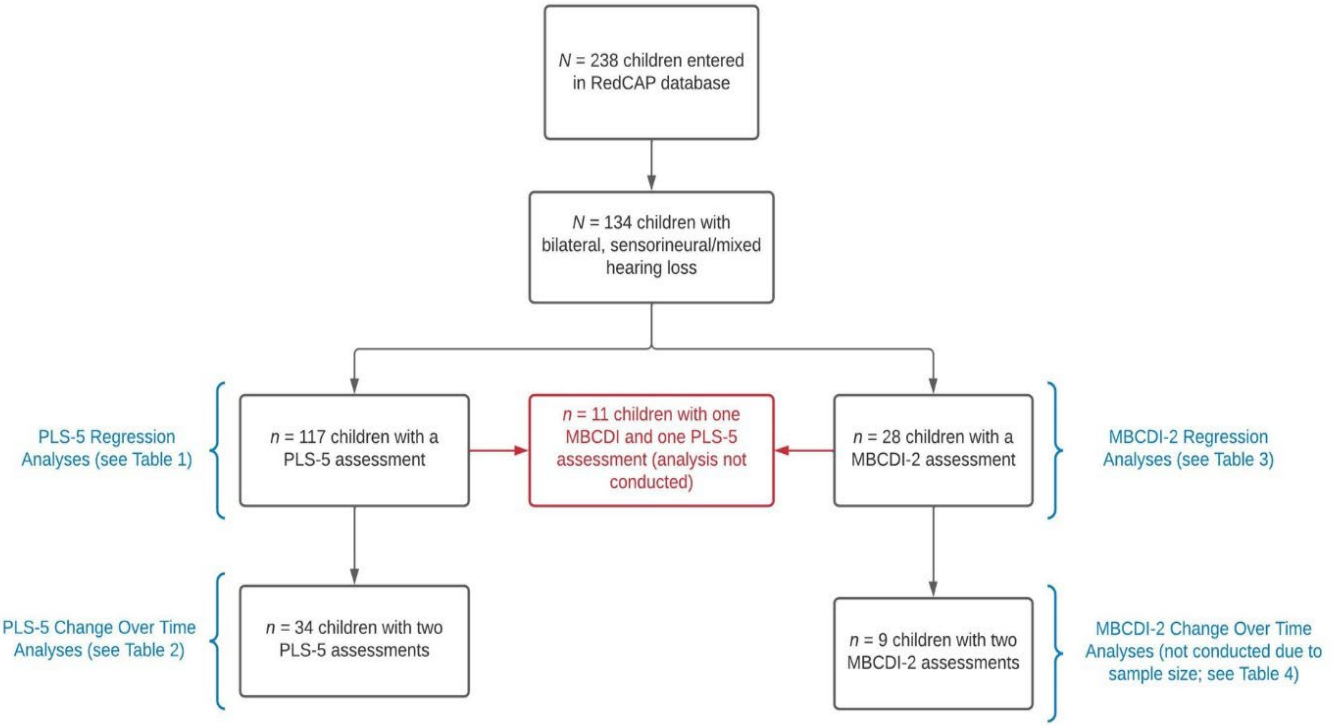
At the end of the pilot, data were available in REDCap for 238 different children. These children had a range of audiological profiles, including unilateral or bilateral, conductive or sensorineural, and ranging from mild to severe in degree. During the pilot study, SLPs were instructed to assess all children enrolled in the IHP at the recommended age ranges and enter the assessment results into the REDCap database. However, we did not have access to the caseload records of the pilot sites, and therefore cannot confirm whether there were children who were DHH for whom SLPs should have conducted an assessment but did not. Therefore, it was unknown whether the children in our database represent all children for whom the procedure should have been used or whether there are groups of children for whom the procedures were not administered.

We can confirm one instance whereby the identification number reported by the SLP could not be linked to an identification number in the program database, and this child was excluded from our analyses. Three children were removed from all analyses for having normal hearing thresholds. In these cases, children were previously under investigation for hearing loss (and so they were assessed by SLPs) but follow-up assessment confirmed normal hearing thresholds.

The analyses for this pilot are based on the entire subset of 134 children who had bilateral sensorineural or mixed hearing losses (see Figure 3). Although the purpose of the Tier 1 outcome monitoring procedure is to document outcomes for all children who receive services from the IHP, very little is known about how unilateral (José et al., 2014) and conductive losses influence spoken language development. There are some data suggesting that children with unilateral losses have poorer spoken language and academic outcomes than children with typical hearing thresholds, although children in these studies tended to be identified later than is the case in the IHP (Fitzpatrick et al., 2019). Similarly, children with conductive losses have a healthy cochlea and their outcomes could reasonably be expected to be different from children with sensorineural losses. Because the primary purpose of this pilot study was to determine whether data generated by the Tier 1 procedure was sensitive to known predictors of spoken language outcomes, we elected to focus our analyses on the groups of children for whom there was the most peer-reviewed data to contextualize our outcomes—children with bilateral sensorineural or mixed hearing losses.

Figure 3

Children from the Overall Sample Included in Pilot Study 1 Analyses



Note. MBCDI = MacArthur-Bates Communicative Development Inventories; PLS = Preschool Language Scale.

After excluding children with normal hearing thresholds, and unilateral and conductive losses, data were available for 117 children with at least one assessment with the PLS-5 (see Table 1) and 34 had data for two assessments (see Table 2). Twenty-eight children had data for at least one assessment with the MBCDI-2 (see Table 3) and nine had data for two assessments (see Table 4). Two children with PLS-5 assessments were fitted with cochlear implants, and 98 were fitted with hearing aids in at least one ear at the time of their language assessment (87 were binaurally fitted, 11 were monaurally fitted). One child with a MBCDI-2 assessment was fitted with a cochlear implant and 19 were fitted with a hearing aid in at least one ear (17 were binaurally fitted, two were monaurally fitted). As a group, children's hearing aids were well-fitted (see Supplemental Materials 1-4 for a comparison of aided Speech Intelligibility Index (SII) to Better Ear Pure Tone Average [BEPTA] to norms reported in Moodie et al., 2017). The decision to fit an ear with a hearing aid is complex and influenced by various factors including the configuration and severity of the child's hearing loss in each ear, and the family's readiness for amplification. Therefore, it is not the case that children in our sample who were not fitted with hearing aids in one, or both ears, *should* have been fitted. Rather, children's audiological profiles at the time of language assessment reflect the family-centered, clinical decision-making of the child's team at the time of their language assessment.

Table 1

Demographics of Children with Data for One PLS-5 Assessment

	Children without Additional Factors (n = 75)		***Children with Additional Factors (n = 41)***	
Variable	*M* (range)	SD	*M* (range)	SD
Age (months)	38.3 (19–71)	7	35.2 (19–71)	12.24
BEPTA (dB HL)	53.2 (17.5–107.5)	23.2	56.49 (26.25–113.33)	22
Better Ear SII (Conversational Speech)	72.5 (5–95)	22.59	68.45 (2–95)	23.88
Better Ear SII (Quiet Speech)	64.29 (2–97)	24.83	66.33 (11–96)	22.1
Expressive Communication (Standard Score)	100.92 (50–150)	20.5	79.67 (53–118)	13.73
Auditory Comprehension (Standard Score)	98.96 (50–137)	19.81	74.49 (50–104)	16.6

Note. BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss; PLS = Preschool Language Scale; SII = Speech Intelligibility Index, the proportion of the speech signal that is audible when the child is wearing their amplification.

Table 2

Demographics of Children with Data for Two PLS-5 Assessments

	Children without Additional Factors (n = 24)		***Children with Additional Factors (n = 9)***	
Variable	*M* (range)	SD	*M* (range)	SD
Age at first PLS-5 (months)	26.96 (19–38)	6.17	28.56 (19–40)	6.1
Age at second PLS-5 (months)	34.76 (24–48)	7.04	34.89 (26–45)	5.8
BEPTA (dB HL)	55.55 (20–107.5)	25.04	67.27 (35–113.33)	27.03
Better Ear SII (Conversational Speech)	69.61 (5–95)	26.13	58.86 (2 – 86)	32.57
Better Ear SII (Quiet Speech)	61.11 (13–97)	26.71	60.33 (11–83)	29.79
First Expressive Communication (Standard Score)	103 (73–123)	14.07	79.63 (68–88)	7.09
Second Expressive Communication (Standard Score)	101.9 (74–122)	14.95	82.75 (72–95)	8.68
First Expressive Communication (Growth Scale Value)	382.25 (297–448)	36.99	328.13 (297–348)	17.73
Second Expressive Communication (Growth Scale Value)	412.5 (314–507)	43.80	362.13 (319–390)	28.22
First Auditory Comprehension (Standard Score)	104.35 (81–127)	13.94	70.88 (53–100)	16.65
Second Auditory Comprehension (Standard Score)	103 (65–123)	14.72	71.75 (54–95)	14.79
First Auditory Comprehension (Growth Scale Value)	394.45 (324–450)	34.27	334.13 (261–392)	41.85
Second Auditory Comprehension (Growth Scale Value)	426.74 (352–504)	36.1	360.63 (304–414)	43.39

Note. BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss; PLS = Preschool Language Scale; SII = Speech Intelligibility Index, the proportion of the speech signal that is audible when the child is wearing their amplification.

Table 3

Demographics of Children with Data for One MBCDI Assessment

	Children without Additional Factors (n = 19)		***Children with Additional Factors (n = 9)***	
Variable	*M* (range)	SD	*M* (range)	SD
Age (months)	12.37 (8–18)	3.14	12.56 (9–18)	2.5
BEPTA (dB HL)	56.23 (31.25–95)	19.77	51.74 (25–95)	24.83
Better Ear SII (Conversational Speech)	72 (21–91)	22.77	60.5 (25–86)	26.29
Better Ear SII (Quiet Speech)	64.17 (6–88)	27.31	67 (56–76)	10.15
Phrases Understood (Percentile Rank)	37.5 (10–75)	19.8	19.11 (< 5–65)	19.89
Words Produced (Percentile Rank)	32.78 (< 5–85)	29.67	21.67 (< 5–45)	16.96
Words Understood (Percentile Rank)	42 (10–99)	25.85	20.22 (< 5–45)	19.26
Gestures (Percentile Rank)	39.67 (< 5–80)	22.61	12.33 (< 5–45)	13.32

Note. BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss; MBCDI = MacArthur-Bates Communicative Development Inventories; SII = Speech Intelligibility Index, the proportion of the speech signal that is audible when the child is wearing their amplification.

Analyses

Data Usability. There were two primary analytic purposes of the Tier 1 pilot. The first was to evaluate whether the Program-level scores (PLS-5 and MBCDI-2) were sensitive to predictors known to influence spoken language outcome in children who are DHH. The second analytic purpose related to data usability was to evaluate whether Program-level scores were sensitive to change for children who had a second assessment using the same test.

The predictors we evaluated for our first purpose included the severity of hearing loss and the presence/absence of additional factors influencing performance. Additional factors were broadly defined as any factor that a SLP believed influenced the child's performance on the test, above and beyond their hearing loss. These additional factors included comorbid diagnoses, social factors such as inconsistent hearing aid use, or children's inability (or unwillingness) to engage in testing. Given the relatively

Table 4

Demographics of Children with Data for Two MBCDI Assessments

	Children without Additional Factors (n = 5)		***Children with Additional Factors (n = 4)***	
Variable	*M* (range)	SD	*M* (range)	SD
Age at first MBCDI (months)	10.5 (8–14)	2.65	11.6 (9–14)	1.95
Age at second MBCDI (months)	15.25 (14–17)	1.26	17.4 (16–19)	1.14
BEPTA (dB HL)	54.5 (31.25–90)	25.6	52.33 (31.67–95)	27.48
Better Ear SII (Conversational Speech)	78 (71–85)	7	41.5 (25–58)	23.33
Better Ear SII (Quiet Speech)	75.67 (64–82)	10.15	69 (69–69)	NA
First Phrases Understood (Percentile Rank)	28.75 (14–45)	13.77	23.6 (< 5–65)	24.99
Second Phrases Understood (Percentile Rank)	28.75 (15–40)	11.09	17.5 (5–40)	15.55
First Words Produced (Percentile Rank)	50 (5–80)	31.88	12 (5–30)	10.95
Second Words Produced (Percentile Rank)	30 (25–40)	7.01	13.75 (5–30)	11.09
First Words Understood (Percentile Rank)	43.75 (20–55)	16.01	23.6 (< 5–45)	20.6
Second Words Understood (Percentile Rank)	30 (10–50)	16.83	8 (< 5–20)	8
First Gestures (Percentile Rank)	36.25 (5–60)	22.23	16.6 (< 5–45)	16.8
Second Gestures (Percentile Rank)	37.5 (15–50)	15.55	13.5 (< 5–20)	7.89

Note. BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss; MBCDI = MacArthur-Bates Communicative Development Inventories; SII = Speech Intelligibility Index, the proportion of the speech signal that is audible when the child is wearing their amplification.

large number of possible factors that could influence language development, it was beyond the scope of the present study to identify which additional factors were differentially associated with language development (e.g., Cupples et al., 2014). For our purposes, we used the presence of additional factors as a coarse indicator that the procedures could be sensitive to additional factors if implementation was scaled up across the province. Follow-up work exploring children's outcomes at the population level would better differentiate the impact of various factors on children's spoken language outcomes.

Prior to analysis, the first author (O.D.) checked the scores recorded in REDCap against the scores reported in the examiner's manuals for the child's recorded chronological age. This process was done to ensure that raw scores were consistently converted to normative scores amongst clinicians, as there is some latitude (particularly with the MBCDI-2) with which to assign percentile ranks. O.D. also checked each child's thresholds from their closest audiology appointment to (but not later than) the Program-level assessment in the IHP database. This was done to determine the child's audiological profile at the time of the language assessment.

Tests' Associations with Predictors. We conducted two direct entry linear regression models to evaluate each test's association with two independent variables (a) severity of hearing loss and (b) the presence/absence of additional factors that SLPs believed may have influenced a child's performance. Severity of hearing loss was conceptualized as the child's Better Ear Pure Tone Average (BEPTA). The presence/absence of additional factors was represented using a dichotomous coding of whether SLPs indicated that they believed factors may have influenced a child's performance as independent variables. Within the IHP, audiometric thresholds must be obtained at 500, 2000, and 4000 Hz in each ear (1000 Hz is discretional; Bagatto et al., 2020; Scollie et al., 2019). Audiologists will attempt to measure all four frequencies in each ear at each assessment, although this may not be possible for various reasons (e.g., child's engagement in testing). Each model's conformity to linear regression assumptions was evaluated using the Global Validation of Linear Models Assumptions, v. 1.0.0.3 in R-Studio (Pena & Slate, 2019).

The first regression model evaluated the association between standard scores for the PLS-5 subtests (auditory comprehension and expressive communication) and the independent variables. The first regression was done using data for a subsample of children who had a PLS-5 assessment. The second regression model evaluated the association between percentile ranks for the MBCDI-2 subtests and the independent variables.

Tests' Sensitivity to Change Over Time. Sensitivity to change over time was coarsely evaluated using paired *t*-tests to compare scores between the first and second assessment intervals. For PLS-5 scores, change was evaluated separately using standard scores and growth scale values, as it has been demonstrated that growth scale values are more sensitive to gains in skills over short intervals (Daub et al., 2017). For the MBCDI-2, change was evaluated using percentile ranks as the test does not report standard scores or growth scale values. We corrected for multiple comparisons using Bonferroni's correction.

Procedure Feasibility. Surveys (see Supplemental Material 5) were designed to identify potential barriers and facilitators to successful implementation. Surveys included 75 questions and asked SLPs to rate their perceptions of the new procedures; their knowledge, skills, and abilities in using the recommended tools; and their opinions on implementation materials and suggestions to improve them. Questions were either in yes/no format or used 5-point Likert scales to measure the strength of SLPs' agreement with statements. Results are reported descriptively.

Results

Data Usability

Tests' Associations with Predictors. All regression analyses met assumptions of normality, independence, homoscedasticity, and linearity with the exception of the PLS-5 Expressive Communication models, which were significantly heteroscedastic. PLS-5 standard scores for both the Auditory Comprehension and Expressive Communication scales were negatively predicted by the presence of additional factors but not BEPTA [auditory comprehension: $F(2, 104) = 21.87$, $p < 0.001$; expressive communication: $F(2,100) = 16.8$, $p < 0.001$; see Table 5]. The combination of BEPTA and the presence of additional factors accounted for 28% and 24% of the variance in children's Auditory Comprehension and Expressive Communication standard scores, respectively (as indicated by R-squared). In both cases, the presence of additional factors was the only significant predictor.

Table 5

Association Between PLS-5 Standard Scores and Predictors

	PLS-5 Standard Score at First Assessment			
	Auditory Comprehension		Expressive Communication	
Predictor	R^2(adj)	b	R^2(adj)	b
Model	0.28*		0.235*	
Better Ear Pure Tone Average (dB HL)		0.263		-0.1
Presence of additional factors affecting outcome		-24.13*		-20.79*

Note. dB HL = decibels Hearing Loss; PLS = Preschool Language Scale.

*$p < 0.001$

The model of the influence of BEPTA and the presence of additional factors on gestures was the only significant model of the MBCDI-2 subtests, $F(2,24) = 5.32$, $p < 0.05$, [phrases understood: $F(2,24) = 2.57$, $p > 0.05$: words produced: $F(2,24) = 0.77$, $p > 0.05$: words understood: $F(2,23) = 2.45$, $p > 0.05$; see Table 6]. Regardless of significance testing, the combination of BEPTA and the presence of additional factors did not explain a large proportion of variance for the phrases understood (11% of variance explained), words produced (-2% of variance explained, indicating exceptionally poor model fit) or words understood (10% of variance explained). The combination of BEPTA and the presence of additional factors accounted for 25% of the variance in children's percentile ranks on the Gestures Produced subtest, although the presence of additional factors was the only significant predictor.

Table 6

Association Between MBCDI-2 Percentile Ranks and Predictors

	MBCDI-2 Percentile Rank at First Assessment							
	Phrases Understood		Words Produced		Words Understood		Gestures	
Predictor	R^2(adj)	b	R^2(adj)	b	R^2(adj)	b	R^2(adj)	b
Model	0.11		-0.02		0.10		0.25*	
Better Ear Pure Tone Average (dB HL)		0.07		-0.17		-0.09		0.01
Presence of additional factors affecting outcome		-17.97*		-12.1		-22.44*		-27.27**

Note. dB HL = decibels Hearing Loss; PLS = Preschool Language Scale.
*$p < 0.001$

Tests' Sensitivity to Change Over Time. With regard to change over time, PLS-5 standard scores did not differ significantly between first [auditory comprehension; $M = 94.26$: expressive communication; $M = 96.04$) and second [auditory comprehension; $M = 93.73$: expressive communication; $M = 96.43$) assessments for either scale [auditory comprehension: $t(26) = 1.5623$, $p > 0.0125$; expressive communication: $t(26) = -0.15823$, $p > 0.0125$]. However, growth scale values were higher at second assessments [auditory comprehension; $M = 405.79$: expressive communication; $M = 396.89$] than first assessments [auditory comprehension; $M = 373.8$: expressive communication; $M = 363,73$) for both subtests [auditory comprehension; $t(26) = 11.623$, $p < 0.0125$: expressive communication; $t(26) = 10.589$, $p < 0.0125$].

We were underpowered to statistically evaluate whether change over time occurred for the MBCDI-2 scores as there were only nine children with data for repeat assessments (see Table 4).

Procedure Feasibility

Fifty-eight SLPs responded to the end of pilot survey, 18 of whom indicated they did not apply the procedure over the one-year pilot. The results for the 40 eligible SLPs are summarized in Appendices A–D. Overall, the majority of SLPs (> 60%) were confident in their knowledge, skills, and abilities to implement the new Program-level outcome monitoring procedures and were confident that they had the physical resources and support from management to do so. There was a lack of strong agreement (< 60%) amongst SLPs that the procedures themselves would be useful within clinical practice and to families. As a group, the majority of SLPs did not agree that the time to administer the Tier 1 procedures either in isolation, or in conjunction with Tier 2 individual vulnerability testing procedures, was appropriate for clinical practice.

Pilot Study 2 – Tier 2 Individual Vulnerability Testing

Method

Procedure

The decision to participate in the Tier 2 individual vulnerability testing pilot during the Tier 1 Program-level pilot was left to the discretion of regional management. Ten of the eleven volunteer sites from Pilot Study 1 agreed to participate in the additional individual vulnerability testing pilot and implement both procedures at the same time. Twenty-three SLPs collected data for the Tier 2 procedure and completed post-pilot surveys to identify barriers and facilitators to implementation.

Participants Assessed in Pilot Study 2

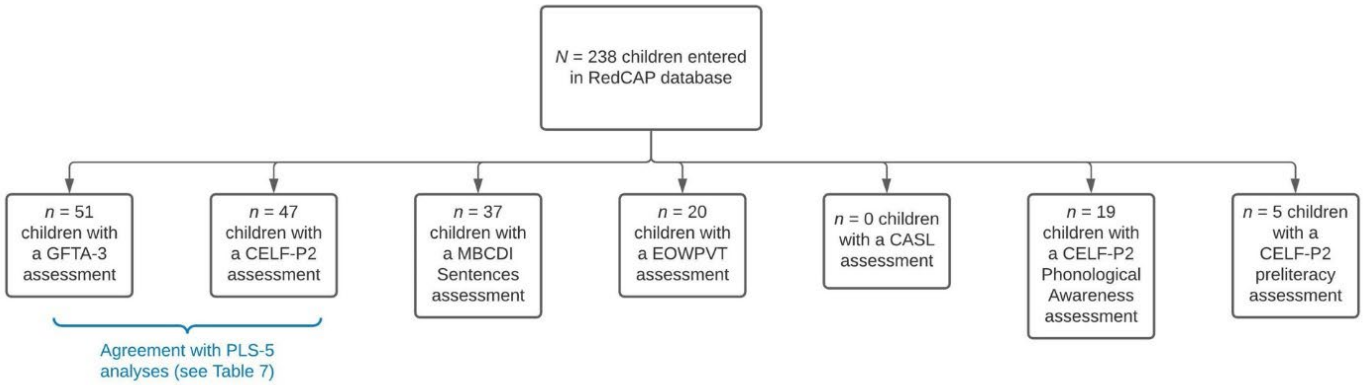
Over the course of the pilot, our team noticed a trend for SLPs from regions that we believed were involved in Pilot Study 2 to flag many children ($n = 72$ of 238) as not being involved in the pilot (i.e., they were only completing assessments from Pilot Study 1). The lack of data for Pilot Study 2 assessments in the REDCap database suggested there was a miscommunication of pilot procedure expectations. It is unclear why this miscommunication occurred, because the decision for a region to participate in Pilot Study 2 was left to regional coordinators. It is possible there was a miscommunication between our research team and the coordinators, between coordinators and SLPs, or a lack of clarity in the implementation materials provided by our team to SLPs.

Once the error was identified, our team reconnected with regional coordinators to confirm their participation and additional communication was provided to SLPs. Following this re-communication, our team observed that SLPs who originally indicated they were not involved in the Tier 2 pilot began to enter individual vulnerability data. However, a significant number of children from these regions were not

assessed with the tests from the Tier 2 procedure before expectations were recommunicated. As a result, a significant amount of expected data (57% of children in piloting regions) was not collected (n = 126 of 238). Reasons for missing data included issues surrounding the original miscommunication (n = 72), and practical limitations (n = 10). Reasons were unknown in 44 cases. Moreover, assessment data for all tests were not reported because the procedure did not require SLPs to administer all tests, but rather gave them choices. The amount of missing data limited our ability to fulfill our primary analytic purposes, but some preliminary hypotheses were developed based on the available data. Our analyses were based on data that were available for children who were assessed using the Goldman-Fristoe Test of Articulation, Third Edition (Goldman & Fristoe, 2015; GFTA-3) and the Clinical Evaluation of Language Fundamentals, Preschool, Second Edition (CELF-P2; Semel et al., 2004; see Figure 4 and Table 7). We included data for all children for whom data were entered in the REDCap database. Children were included regardless of audiological profile (i.e., we included children with unilateral and conductive losses), as our primary aim was to explore whether the Tier 2 tests agreed in their characterization of whether a child had an impairment based on Tier 1 testing regardless of hearing characteristics.

Figure 4

Children from the Overall Sample Included in Pilot Study 2 Analyses



Note. GFTA-3 = Goldman-Fristoe Test of Articulation, 3rd ed.; CELF-P2 = Clinical Evaluation of Language Fundamentals, Preschool, 2nd ed.; MBCDI = MacArthur-Bates Communicative Development Inventories; EOWPVT = Expressive One Word Picture Vocabulary Test; PLS-5 = Preschool Language Scale, 5th ed. CASL-2 = Comprehensive Assessment of Language Fundamentals, 2nd ed.

Table 7

Demographics of Children Included in Pilot Study 2

	Children with GFTA-3 Sounds-in-Words Assessments (n = 48)			***Children with CELF-P2 Word Structure Assessments (n = 46)***		
Variable	*n*	*M* (range)	SD	*n*	*M* (range)	SD
Presence of Additional Factors	16			10		
Bilateral Hearing Loss	31			34		
BEPTA (dB HL)		51.47 (18.75–98.33)*	24.17		47.43 (17.25–92.5)*	23.18
Unilateral Hearing Loss	17			12		
PTA (dB HL)		47.74 (26.25–81.25)	18.84		47.96 (28.75–83.75)	16.6
Conductive Hearing Loss	4			4		
Age at PLS Assessment (months)		43.54 (31–71)	8.58		48.63 (34–71)	9.48
Expressive Communication (Standard Score)		103.85 (64–150)	19.12		103.53 (54–150)	22.14
Auditory Comprehension (Standard Score)		103.34 (73–150)	16.67		101.55 (51–150)	20.26
Age at IVT Assessment (months)		44 (31–71)	6.68		49.19 (37–72)	9.27
GFTA-3 Sounds-in-Words (Standard Score)		89.32 (42–123)	18.2			
CELF-P2 Word Structure (Percentile Ranks*)					42.2 (0.1-99)	35.74

Note. GFTA-3 = Goldman-Fristoe Test of Articulation, 3rd Ed.; CELF-P2 = Clinical Evaluation of Language Fundamentals, Preschool – 2nd Ed.; BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss; PLS = Preschool Language Scale; PTA = pure-tone average; IVT = Individual Vulnerability Test. Pure tone averages < 25 db were the result of high or low frequency hearing losses, where the child experienced hearing losses at some, but not all, frequencies. Standard scores are not available for the CELF-P2 subtests.

Analysis

Data Usability

We had intended to develop a structural equation model to map the relations between overall spoken language assessment data (collected using the Tier 1 PLS-5 and MBCDI-2) and different domains of spoken language (data collected using the Tier 2 individual vulnerability testing procedures), but we were underpowered to perform these analyses because of the large amount of missing data. Instead, we explored the level of classification agreement (whether a child performed below age expectations on a given test) between the two most frequently completed tests (the GFTA-3 Sounds-in-Words subtest, which measures accuracy of articulation of consonants and consonant clusters during single word production, and the CELF-P2 Word Structure subtest, which measures accuracy of grammar use) and scores on the PLS-5. As in Pilot Study 1, the first author (O.D.) corrected the GFTA-3 standard scores and CELF-P2 percentiles to ensure consistency and developed audiological profiles for each child. For both subtests, few corrections were required (14% for the GFTA-3, and 2% for the CELF-P2).

Children's scores for the GFTA-3 Sounds-in-Words and CELF-P2 Word Structure subtests and both scales of the PLS-5 were categorized as *within*, *borderline*, or *below* age expectations based on the cut-score with the highest diagnostic accuracy as well as a 90% confidence interval around that score. For instance, the PLS-5 has the highest diagnostic accuracy when a cut-score of -1*SD* below the mean is used. In this case, we categorized a score as *below* age-expectations when the child's standard score was lower than the cut-score *and* the upper-bound of the 90% confidence interval was also below the cut-score. Similarly, a score was considered *within* age expectations when the child's score was above the cut-score and the lower bound of the 90% confidence interval was above the cut-score. In cases where the child's score was diagnostically ambiguous (the upper and lower bound of the confidence interval fell both above and below the cut-score), the child was categorized as *Borderline*. Instances were then tallied when the categorization of the PLS-5 was higher, the same as, or lower than the diagnostic categorization of the companion tests. Finally, we evaluated agreement between the PLS-5 auditory comprehension and expressive communication scales and GFTA-3 and CELF-P2 subtests using a Kendall's Coefficient of Concordance.

Procedure Feasibility

Survey data were analyzed descriptively as in Pilot Study 1.

Results

Agreement between Tier 1 and Tier 2 Tests in Diagnostic Categorizations

The proportions of children considered *within, borderline,* or *below* age expectations for each test are reported in Tables 8–11. Children's categorization on both PLS-5 auditory comprehension and expressive communication scales agreed with one another, and with diagnostic categorization on the GFTA-3 ($Wt(46) = 0.71$, $p < 0.05$) and CELF-P2 ($Wt(43) = 0.73$, $p < 0.05$). Analyses were not repeated for scores on the other tests included in the Tier 2 procedure because of the small amount of data available for each assessment and a lack of sensitivity/specificity data to define within/borderline/below age expectations for the MBCDI Words and Sentences form.

Table 8

Agreement Between PLS-5 Auditory Comprehension and GFTA-3 Sounds-in-Words Subtest

	GFTA-3		
PLS-5	**Within**	**Border**	**Below**
Within	32	4	1
Border	4	0	2
Below	0	0	5

Note. PLS = Preschool Language Scale; GFTA-3 = Goldman-Fristoe Test of Articulation, 3rd ed.

Table 9

Agreement Between PLS-5 Expressive Communication and GFTA-3 Sounds-in-Words Subtest

	GFTA-3		
PLS-5	**Within**	**Border**	**Below**
Within	31	3	2
Border	4	1	4
Below	0	0	2

Note. PLS = Preschool Language Scale; GFTA-3 = Goldman-Fristoe Test of Articulation, 3rd ed.

Table 10

Agreement between PLS-5 Auditory Comprehension and CELF-P2 Words Structure Subtest

	CELF-P2		
PLS-5	**Within**	**Border**	**Below**
Within	17	12	4
Border	0	5	2
Below	0	0	5

Note. PLS = Preschool Language Scale; CELF-P2 = Clinical Evaluation of Language Fundamentals, Preschool, 2nd ed.

Table 11

Agreement Between PLS-5 Expressive Communication and CELF-P2 Words Structure Subtest

	CELF-P2		
PLS-5	**Within**	**Border**	**Below**
Within	16	12	3
Border	1	5	2
Below	0	0	5

Note. PLS = Preschool Language Scale; CELF-P2 = Clinical Evaluation of Language Fundamentals, Preschool, 2nd ed.

Procedure Feasibility

At the end of the pilot, 36 SLPs completed online surveys to provide feedback on the new Tier 2 procedures. Thirteen SLPs indicated that they did not use the individual vulnerability testing procedure at all over the course of the pilot, and therefore did not complete the remaining survey questions. Summaries of the remaining 23 SLPs' responses are outlined in Appendices E–I.

As was the case with the Tier 1 Program-level outcome monitoring procedures, the majority (> 60%) of SLPs were confident in their knowledge, skills, and abilities to implement the Tier 2 individual vulnerability testing procedures, with the notable exception of the *Comprehensive Assessment of Language Fundamentals*, *2nd edition* (CASL-2; Carrow-Woodfolk, 2017). The majority of SLPs also reported that they had resources such as test manuals (except for the CASL-2) and managerial support. Most SLPs agreed or strongly agreed that results from the Tier 2 testing supported their clinical decision making and could be used to improve services for families of children who are DHH.

Although SLPs reported that the individual vulnerability test process provided valuable information, there was a lack of consensus about whether the amount of time required to implement was feasible. The percentage of SLPs who reported being able to consistently implement the Tier 2 process was also divided, and 78% of respondents reported that additional administrative support or time release from other clinical duties would be helpful for implementing it. In open-ended comments, some SLPs reported concerns that the combination of Tier 1 and Tier 2 testing was overly burdensome for children, families, and themselves. Finally, when asked whether it would be helpful to forgo Tier 2 testing altogether, 47% of SLPs reported feeling neutral, and the remaining SLPs were divided between agreeing and disagreeing.

Discussion

These two pilot studies present preliminary evidence for the usability and feasibility of the spoken language outcome monitoring procedure developed by Daub and Oram Cardy (2021). For program evaluation purposes, repeated assessment using a narrow set of omnibus language tests (i.e., the MBCDI-2 and the PLS-5) was expected to support group level analysis of outcomes for children who are DHH. By using the same tests over time, we expected that any changes we observed would be attributable to the child's development, rather than changes in the psychometric properties of the assessment tools. This is the first account, to our knowledge, of an effort to evaluate a spoken language outcome monitoring procedure for an EHDI program. Although the need for routine spoken language outcome monitoring is clear (JCIH, 2007, 2013, 2019; Moeller et al., 2013), there is limited guidance for how to accomplish the diverse assessment purposes proposed under these recommendations.

Data Usability

Data from Pilot Study 1 suggest that the PLS-5 might be appropriate for fulfilling program evaluation purposes, however there was less evidence to support use of the MBCDI-2. PLS-5 growth scale values were sensitive to change over time (Daub et al., 2017) and standard scores were predicted by additional factors, so in this regard, the PLS-5 conformed to our prediction that it would capture growth in children's spoken language skills. The MBCDI-2 did not conform to our prediction, although we did not have a large enough sample of children with two MBCDI-2 assessments to adequately evaluate whether the MBCDI-2 scores changed over time.

There are several possible explanations for the lack of evidence to support using the MBCDI-2. First, it is possible that the impact of hearing loss on the aspects of language measured by the MBCDI-2 Words and Gestures form is not observed in very young children ($M < 14$ months, in our sample). Without data to compare performance on the PLS-5 in children under 18 months, we cannot be assured that the PLS-5 would have been any more informative at this young age. Our findings might also be explained by the scoring characteristics of the MBCDI-2 itself: it has been well documented that there is a wide range of typical variation associated with MBCDI-2 scores, particularly with regard to words produced in children younger than 18 months (Fenson et al., 2000; Feldman et al., 2000). Further, a single total number of words can correspond to a wide range of percentile ranks and small changes in total scores can dramatically influence a child's percentile rank. For example, for an 8-month-old boy who produces no words, a percentile rank of between 5 and 55 can be assigned, whereas an 8-month-old boy producing a single word corresponds to a percentile rank of either 65 or 70 (Fenson et al., 2007, p. 120). Therefore, the scoring properties of the MBCDI-2 may mean that it is not sensitive enough to use as a Program-level outcome measure in young children.

Why neither test was predicted by severity of hearing loss (BEPTA) is less immediately clear. The lack of an effect is particularly surprising for the PLS-5 for several reasons. First, the use of standard scores rather than percentile ranks allows for more precise scoring than the MBCDI-2. Second, we used the PLS-5 for a much broader age range than the MBCDI-2 and the lack of effect cannot be accounted for by the age of the children in our sample. We also had a much larger sample for the PLS-5 analyses than the MBCDI-2 and the lack of effect cannot be explained by a lack of power. Finally, we had a wide range of both PLS-5 scores (e.g., between 50 and 150) and BEPTA (e.g., 20–107.5). For both variables, we had data representing the full range of possible values and our null finding cannot be accounted for by range restriction of either variable. Interestingly, the average PLS-5 scores in our sample (for children without additional factors) were higher than what is typically reported in outcome studies (e.g., Tomblin et al., 2015) and approximate a normal distribution, which has a mean standard score of

100 and standard deviation of 15. In our data, children without additional factors (as a group) had a mean of 100.92 (*SD* = 20.5) on the expressive communication scale and a mean of 98.98 (*SD* = 19.81) on the auditory comprehension scale (see Supplemental Material 5). This raises the possibility that perhaps the lack of influence of BEPTA on PLS-5 scores accurately reflects children's spoken language outcomes. All children in our sample were receiving comprehensive EHDI services and wearing well-fitted hearing aids (see Supplemental Materials 1–4). If an EHDI program's goal is to support age-appropriate language outcomes by providing children with consistent access to auditory information, then it is reasonable to expect that severity of hearing loss should not predict outcomes but other variables (e.g., additional factors influencing performance) would. In our data, additional factors were broadly defined as any factor SLPs believed may influence a child's performance on the test, above and beyond their hearing loss. Once those factors were statistically controlled for (by entering the variable into our regressions), severity of hearing loss did not uniquely contribute to children's performance.

It may be the case that our data is preliminary evidence that the IHP is achieving their goal of ameliorating the impact of inconsistent auditory access on spoken language outcomes. That is, perhaps the impact of severity of hearing loss on spoken language development is mitigated by response to intervention. This idea is consistent with previous work suggesting that children with lower language skills and more severe hearing loss show greater gains in PLS-5 growth scale values after amplification (Daub et al., 2017). This idea also aligns with research showing children with permanent hearing loss catch up to their peers with typical hearing thresholds over time as a function of access to auditory information (conceptualized as consistent hearing aid use and quality of hearing aid fit; Tomblin et al., 2015). However, we remain cautious in our interpretation of the data. Without access to SLPs' caseloads to ensure that *all* children in the IHP were reflected in our data, we cannot confirm that our sample is representative of the IHP. Future work using population-level data from the IHP will model how children's spoken language outcomes change over time, and as a function of intervention characteristics such as quality of hearing aid fit. We are also cautious in our interpretations because our sample was insufficient to identify whether some additional factors differentially interacted with severity of hearing loss in predicting spoken language outcomes. There is some evidence that certain comorbid diagnoses (e.g., autism, cerebral palsy, developmental delay) are particularly influential in spoken language outcomes of children who are DHH (Cupples et al., 2014). It is also possible that some performance factors (e.g., inattention during testing) influenced children's hearing thresholds. Future work evaluating the outcomes of children across the entire IHP is warranted to identify whether the lack of effect of BEPTA on children's spoken language outcomes holds for children with, and without, additional complicating factors.

Pilot Study 2 was conducted to evaluate the usability of data from an individual vulnerability testing procedure. Because children who are DHH have ongoing inconsistent access to auditory information, it has been documented that they continue to struggle in certain domains of spoken language (e.g., Moeller et al., 2007) even when they may perform within normal limits on omnibus measures. As a result, an outcome monitoring procedure that only reports on spoken language outcomes broadly has the potential to over-estimate children's abilities and miss opportunities to develop additional supports for specific domains of spoken language development. Due to missing data, we were unable to fulfill our planned analyses, however, preliminary analyses exploring the agreement between overall language comprehension and use of language (PLS-5) with articulation (GFTA-3) and grammar (CELF-P2) indicated that diagnostic categorizations largely agreed. Our data were insufficient to report on whether the individual vulnerability testing procedure provided unique clinical information. Note that these analyses do not account for all domains of language that we planned to measure, nor do they account for longitudinal relationships between measures. Future, longitudinal research evaluating this procedure on a larger and more representative sample of children who are DHH is needed to draw definitive theoretical and clinical conclusions.

Procedure Feasibility

Both pilot studies evaluated the feasibility of the recommended procedures through a descriptive evaluation of SLPs' survey responses. For both the Tier 1 Program-level outcome monitoring procedure and the Tier 2 individual vulnerability testing procedure, SLPs reported a high degree of confidence in their knowledge and skills to implement the procedures accurately. In both pilot studies, SLPs flagged concerns about the amount of time it took to complete the procedures. Note that most SLPs participating in Pilot Study 1 were also participating in Pilot Study 2. Therefore, we are unable to identify whether SLPs' perceptions of the amount of time each procedure took was a true reflection of each procedure independently *or* if completing both procedures simultaneously impacted their perceptions.

The key difference in SLPs' perceptions between the two pilot studies related to clinical relevance. As a group, SLPs were less convinced of the value of the Tier 1 Program-level outcome monitoring procedure than they were of the Tier 2 individual vulnerability testing procedure. Although we are cautious in the generalizability of this finding because of the small number of SLPs who completed surveys in Pilot Study 2, it is not necessarily surprising. The Tier 1 Program-level outcome monitoring procedure was intended to support program evaluation and we know that many children who are DHH perform within normal limits on omnibus language assessments but still have needs in certain domains of language. Although our usability data for the individual vulnerability testing pilot was insufficient to make recommendations for EHDI programs and to determine whether tests provided unique

predictive information, SLPs' feedback indicates that valuable clinical information may be gained from the Tier 2 procedure. Future work is warranted where administration of tests can be more closely controlled to evaluate the relation between the proposed measures in the individual vulnerability testing procedure.

Limitations & Future Directions

Naturally, the results of our feasibility analyses are heavily dependent on the IHP's context and may not necessarily generalize to other EHDI programs. However, our results provide preliminary evidence that the procedures recommended in Daub and Oram Cardy (2021) are possible to implement, and are largely perceived as informative by SLPs. In addition to the findings reported here, our surveys (see Supplemental Materials) can support other EHDI programs in evaluating their own procedures.

Finally, it is unknown whether the procedures we evaluated are implementable at the scale of an entire EHDI program, whether appropriate implementation is sustainable over time and survives staff turnover, and whether the data collected here can be used to benefit programs, families, and children who are DHH. Future work will monitor use of the procedures over time and document the impact of data on program planning and services.

Taken together, results highlight the importance of carefully considering the questions EHDI programs seek to answer with spoken language outcome monitoring and the methods they use to answer these questions. Testing is not a neutral activity. There are costs associated with engaging in testing including using limited resources to test rather than allocating those resources elsewhere (e.g., intervention). There are also costs for children and their families who engage in testing such as time and emotional impact of engaging in repeated testing (e.g., frustration with their child's progress; Daub et al., 2021; Messick, 1993). Risks associated with testing for families and children who are DHH are another factor that must be considered. If inappropriate tests are used, or data are misinterpreted, SLPs may draw erroneous conclusions about the effectiveness of an intervention, or about children who are DHH themselves. If the data that are collected during spoken language outcome monitoring cannot answer the questions they were intended to, then the costs and risks are not justified. If the procedure used to collect data is too burdensome to be implemented consistently and accurately, then the resulting data may become unusable and testing is similarly unjustified. The data reported here suggest that our proposed Tier 1 Program-level procedure may result in data appropriate for our intended purposes, but we have insufficient evidence to justify the implementation of the Tier 2 individual vulnerability testing procedure in clinical practice. In presenting these findings to the IHP, we recommended adoption of the Program-level procedure as originally defined with regular data monitoring for the first two years to verify whether the data are suitable at the scale of the entire program. For the individual vulnerability testing, we recommended sharing the tests we selected for Tier 2 monitoring with SLPs (Daub & Oram Cardy, 2021), and the rationale for monitoring key areas of vulnerability in children who are DHH. This would leave SLPs free to use the recommended Tier 2 tools when they identify a need in clinical practice, rather than mandating it program-wide.

Although spoken language outcome monitoring is predicted to support various stakeholders' decision-making (JCIH, 2007, 2013, 2019), if spoken language outcome monitoring procedures fail to improve programs or children's outcomes in practice, then the efforts spent regularly assessing children's spoken language development might be better spent elsewhere. As interdisciplinary professionals invested in improving outcomes for children who are DHH, it is imperative that we grapple with these psychometric and implementation issues in the design and evaluation of EHDI programs.

Conclusions

This paper summarizes preliminary evidence of the usability and feasibility of a spoken language outcome monitoring procedure for EHDI programs. This evidence suggests that the Tier 1 Program-level procedure may be feasible to implement and result in usable data, although future work is needed to evaluate whether the data are sufficient to address program evaluation needs once implemented across the IHP. There was insufficient evidence to recommend the use of the Tier 2 individual vulnerability testing procedures to implement in EHDI programs at this point. Future work will evaluate whether the procedure can be accurately implemented, whether accurate implementation can be sustained over time, and whether the procedure influences decision-making to improve program and children's outcomes.

References

Bagatto, M., Easwar, V., El-Naji, R., Hyde, M., Malandrino, A., Martin, V., Pigeon, M., Purcell, D., Scollie, S., & Witte, J. (2020). Protocol for Auditory Brainstem Response – Based Audiological Assessment (ABRA) Version 2018.02. Developed for the *Ontario Ministry of Children, Community and Social Services' Infant Hearing Program*; Retrieved from: https://www.uwo.ca/nca/pdfs/clinical_protocols/2018.01%20ABRA%20Protocol_Oct%2031.pdf

Canadian Infant Hearing Task Force. (2014). Report card of early hearing detection and intervention. Retrieved from http://sacoac.ca/sites/default/files/resources/Report%20Card-2014_EN.pdf

Canadian Infant Hearing Task Force. (2019). 2019 Report card on Canadian Early Hearing Detection and Intervention Programs. Retrieved from https://canadianaudiology.ca/wp-content/uploads/2019/04/Report-Card-2019_FINAL_VERSION_EN.pdf

Carrow-Woolfolk, E. (2017). *Comprehensive Assessment of Spoken Language, 2nd ed.* San Antonio, TX: Pearson Assessments.

Ching, T. Y. C., Dillon, H., Button, L., Seeto, M., Van Buynder, P., Marnane, V., Cupples, L., & Leigh, G. (2017). Age at intervention for permanent hearing loss and 5-year language outcomes. *Pediatrics*, *140*(3), e20164274. https://doi.org/10.1542/peds.2016-4274

Cunningham, B. J., Daub, O., & Oram Cardy, J. (2021). Implementing evidence-based assessment practices for the monitoring of spoken language outcomes in children who are deaf/hard of hearing in a large-scale community program. *Canadian Journal of Speech-Language Pathology and Audiology*, *45*, 41–58. https://www.cjslpa.ca/detail.php?ID=1272&lang=en

Cupples, L., Ching, T.Y.C., Crowe, K., Seeto, M., Leigh, G., Street, L., Day, J., Marnane, V., & Thomson, J. (2014). Outcomes of 3-year-old children with hearing loss and different types of additional disabilities. *The Journal of Deaf Studies and Deaf Education, 19*(1), 20–39. https://doi.org/10.1093/deafed/ent039

Daub, O., Bagatto, M. P., Johnson, A. M., & Oram Cardy, J. (2017). Language outcomes in children who are deaf and hard of hearing: The role of language ability before hearing aid intervention. *Journal of Speech, Language, and Hearing Research*, *60*(11), 3310–3320. https://doi.org/10.1044/2017_JSLHR-L-16-0222

Daub, O., Cunningham, B. J., Bagatto, M. P., Johnson, A. M., Kwok, E. Y. L., Smyth, R. E., & Oram Cardy, J. (2021). Adopting a conceptual validity framework for testing in speech-language pathology. *American Journal of Speech-Language Pathology, 30,* 1894–1908. https://doi.org/10.1044/2021_AJSLP-20-00032

Daub, O., & Oram Cardy, J. (2021). Developing a spoken language outcome monitoring procedure for a Canadian Early Hearing Detection and Intervention program: Process and recommendations. *Journal of Early Hearing Detection and Intervention, 6*(1), 12–31. https://doi.org/10.26077/3208-8406

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). Reply: 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., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *The MacArthur-Bates Communicative Development Inventories User's Guide and Technical Manual 2nd edition.* Brookes Publishing.

Fitzpatrick, E. M., Gaboury, I., Durieux-Smith, A., Coyle, D., Whittingham, J., & Nassrallah, F. (2019). Auditory and language outcomes in children with unilateral hearing loss. *Hearing Research*, *372*, 42–51. https://doi.org/10.1016/j.heares.2018.03.015

Goldman, R., & Fristoe, M. (2015). *Goldman-Fristoe Test of Articulation, 3rd ed.* Pearson Education Inc.

Graham, I. D., & Logan, J. (2004). Innovations in knowledge transfer and continuity of care. *Canadian Journal of Nursing Research 36*(2), 89–103.

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.

Joint Committee on Infant Hearing. (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.

Joint Committee on Infant Hearing. (2019). Year 2019 position statement: Principles and guidelines for early hearing detection and intervention programs. *Journal of Early Hearing Detection and Intervention, 4*(2), 1–44. https://doi.org/10.15142/fptk-b748

José, M. R., Mondelli, M. F., Feniman, M. R., & Lopes-Herrera, S. A. (2014). Language disorders in children with unilateral hearing loss: A systematic review. *International Archives of Otorhinolaryngology, 18*, 198–203. https://doi.org/10.1055/s-0033-1358580

Messick, S. (1993). Foundations of validity: Meaning and consequences of psychological assessment. *Educational Testing Service Research Report Series, 2.* *https://doi.org/10.1002/j.2333-8504.1993.tb01562.x*

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., & Tomblin, J. B. (2015). An introduction to the Outcomes of Children with Hearing Loss study. *Ear and Hearing, 36*, 4S–13S. https://doi.org/10.1097/AUD.0000000000000210

Moeller, M. P., Tomblin, J. B., Yoshinaga-Itano, C., Connor, C. M., & Jerger, S. (2007). Current state of knowledge: Language and literacy of children with hearing impairment. *Ear and Hearing*, *28*(6), 740–753. https://doi.org/10.1097/AUD.0b013e318157f07f

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 v.1.0). *Trends in Amplification, 15*(1-2), 34–56. https://doi.org/10.1177/1084713811417634

Moodie, S. T. F., The Network of Pediatric Audiologists of Canada, Scollie, S. D., Bagatto, M. P., & Keene, K. (2017). Fit-to-targets for the desired sensation level version 5.0a hearing aid prescription method for children. *American Journal of Audiology, 26*(3), 251–268. https://doi.org/10.1044/2017_AJA-16-0054

Pena, E. A., & Slate, E. H. (2019). Global validation of linear models assumptions 1.0.0.3. CRAN. https://cran.r-project.org/web/packages/gvlma/gvlma.pdf

Scollie, S., Pigeon, M., Bagatto, M., Witte, J., & Malandrino, A. (2019). Audiometric assessment for children aged 6 to 60 months, Version 2019.01. Developed for the *Ontario Ministry of Children, Community and Social Services' Infant Hearing Program*. https://www.uwo.ca/nca/pdfs/clinical_protocols/IHP_CBA%20Protocol_2019.01.pdf

Semel, E., Wiig, E., & Secord, W. A. (2004). *Clinical Evaluation of Language Fundamentals – Preschool, 2nd ed.* Pearson Education Inc.

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/10.1097/AUD.0000000000000219

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale, 4th edition*. Pearson Education Inc.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool Language Scale, 5th edition*. Pearson Education Inc.

EHDInfo

IOWA EHDI Program recognized as the best website among the website of the year awards for the past 12 years.

Appendix A

Speech-Language Pathologists' Opinions on the Tier 1 Procedure Statement

Statement	Strongly Disagree *n* (%)	Disagree *n* (%)	Neither agree nor disagree *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (Range)	Not applicable *n* (%)
The IHP's new Program-level Outcome Monitoring Procedure was useful for my clinical practice.	1 (2.5%)	4 (10%)	17 (42.5%)	15 (37.5%)	3 (7.5%%)	3 (1–5)	0 (0%)
I was able to consistently implement the new Program-level Outcome Monitoring recommendations in my practice.	1 (2.5%)	14 (35%)	7 (17.5%)	17 (42.5%)	1 (2.5%)	4 (1–5)	0 (0%)

Note. IHP = Infant Hearing Program.

Appendix B

Speech-Language Pathologists' (SLP) Opinions of their Capacity to Implement the Tier 1 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neither agree nor disagree *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (Range)	Not answered *n* (%)
Over the past year I felt I had the clinical skills required to implement the new Program-level Outcome Monitoring Procedures.	0 (0%)	0 (0%)	3 (7.5%)	26 (65%)	11 (27.5%)	4 (3–5)	0 (0%)
I am familiar with the administration of the MacArthur-Bates Communicative Development Inventories- Words & Gestures (MBCDI).	1 (2.5%)	1 (2.5%)	2 (5%)	25 (62.5%)	10 (25%)	4 (1–5)	1 (2.5%)
I was able to accurately score and use the norms tables for the MacArthur-Bates Communicative Development Inventories Words & Gestures (MBCDI).	2 (5%)	1 (2.5%)	5 (12)	19 (47.5%)	9 (22.5%)	4 (1–5)	4 (10%)
I am familiar with the administration of the Preschool Language Scales-5th Edition.	0 (0%)	0 (0%)	0 (0%)	21 (52.5%)	19 (47.5%)	4 (4–5)	0 (0%)
I was able to accurately score and use the norms tables for the Preschool Language Scales-5th Edition.	0 (0%)	0 (0%)	0 (0%)	18 (45%)	21 (52.5%)	5 (4–5)	1 (2.5%)
The new Program-level Outcome Monitoring Procedures have helped me with my clinical decision-making.	3 (7.5%)	7 (17.5%)	12 (30%)	14 (35%)	3 (7.5%)	4 (1–5)	1 (2.5%)
The new Program-level Outcome Monitoring Procedures have helped parents with their decision-making.	1 (2.5%)	4 (10%)	21 (52.5%)	9 (22.5%)	3 (7.5%)	3 (1–5)	2 (5%)
Repeat administration of the Program-level Outcome Monitoring tools to the same child 6–12 months later benefited the families and children that I serve.	1 (2.5%)	3 (7.5%)	15 (37.5%)	12 (30%)	4 (10%)	3 (1–5)	5 (12.5%)
Repeat administration of the Program-level Outcome Monitoring tools to the same child was useful for my own clinical practice.	1 (2.5%)	3 (7.5%)	11 (27.5%)	18 (45%)	3 (7.5%)	4 (1–5)	4 (10%)

Appendix C

Speech-Language Pathologists' (SLP) Opinions on the Practice Environment and the Tier 1 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neither agree nor disagree *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (Range)	Not answered *n* (%)
The length of time it took to administer the recommended Program-level Outcome Monitoring tests was appropriate for incorporation into routine clinical practice.	5 (12.5%)	10 (25%)	7 (17.5%)	15 (37.5%)	2 (5%)	4 (1–5)	1 (2.5%)
The length of time it took to score and interpret the results of the recommended Program-level Outcome Monitoring tests was appropriate for incorporating into routine clinical practice.	3 (7.5%)	4 (10%)	13 (32.5%)	19 (47.5%)	2 (5%)	4 (1–5)	1 (2.5%)
The length of time it took to talk with parents about results of the recommended Program-level Outcome Monitoring tests was appropriate for incorporation into clinical practice.	1 (2.5%)	4 (10%)	13 (32.5%)	19 (47.5%)	2 (5%)	4 (1–5)	1 (2.5%)
The time it took to do the recommended Program-level Outcome Monitoring and reporting did NOT negatively impact other areas of my practice.	2 (5%)	9 (22.5%)	12 (30%)	14 (35%)	2 (5%)	4 (1–5)	1 (2.5%)
The environment in which I worked made it difficult for me to implement the recommended Program-level Outcome Monitoring.	7 (17.5%)	15 (37.5%)	8 (20%)	7 (17.5%)	0 (0%)	2 (1–4)	3 (7.5%)
I had the supplies I needed (e.g., test forms) to implement the new Program-level Outcome Monitoring.	0 (0%)	2 (5%)	0 (0%)	19 (47.5%)	19 (47.5)	5 (2–5)	0 (0%)
When I had a question about the Program-level Outcome Monitoring Procedures, I consulted with my colleagues.	0 (0%)	3 (7.5%)	4 (10%)	24 (60%)	5 (12.5%)	4 (2–5)	4 (10%)
When I had a question about the Program-level Outcome Monitoring Procedures, I consulted with my managers/ administrators.	1 (2.5%)	8 (20%)	7 (17.5%)	16 (40%)	1 (2.5%)	4 (1–5)	7 (17.5%)
When I had a question about the Program-level Outcome Monitoring Procedures, I consulted the "Pilot Implementation Q&A" section of Western's OWL site.	1 (2.5%)	8 (20%)	4 (10%)	20 (50%)	4 (10%)	4 (1–5)	3 (7.5%)

Appendix C (cont.)

Speech-Language Pathologists' (SLP) Opinions on the Practice Environment and the Tier 1 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neither agree nor disagree *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (Range)	Not answered *n* (%)
I had the resources I needed (e.g., administrative support for scheduling, data entry) to do the new Program-level Outcome Monitoring Procedures.	4 (10%)	13 (32.5%)	8 (20%)	11 (27.5%)	3 (7.5%)	2 (1–5)	1 (2.5%)
I had permission from my manager to take the time I needed to complete Program-level Outcome Monitoring Procedures.	0 (0%)	0 (0%)	5 (12.5%)	27 (67.5%)	6 (15%)	4 (3–5)	2 (5%)
Getting timely feedback from experts (i.e., the research team at Western University) helped me to implement the new Program-level Outcome Monitoring Procedures.	0 (0%)	2 (5%)	15 (37.5%)	13 (32.5%)	5 (32.5%)	3 (2–5)	5 (12.5%)
The SLPs I worked with were excited about the new Program-level Outcome Monitoring Procedures.	5 (12.5%)	8 (20%)	18 (45%)	4 (10%)	2 (5%)	3 (1–5)	3 (7.5%)
Managers/administrators I worked with were supportive of the new Program-level Outcome Monitoring Procedures.	0 (0%)	0 (0%)	10 (25%)	24 (60%)	4 (10%)	4 (3–5)	2 (5%)
The parents I worked with were interested in the results of the new Program-level Outcome Monitoring Procedures.	5 (12.5%)	2 (5%)	21 (52.5)	10 (25%)	1 (2.5)	3 (1–5)	1 (2.5%)
The task of completing the MBCDI was not too difficult for parents (respondents) to perform.	5 (12.5%)	3 (7.5%)	7 (17.5%)	19 (47.5%)	3 (7.5%)	4 (1–5)	3 (7.5%)
The task of completing the MBCDI was not too time consuming for parents (respondents) to perform.	3 (7.5%)	9 (22.5%)	9 (22.5%)	15 (37.5%)	1 (2.5%)	4 (1–5)	3 (7.5%)

Note. MBCDI = MacArthur-Bates Communicative Development Inventories.

Appendix D

Speech Language Pathologists' (SLP) Opinions on the Quality of the Tier 1 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neither agree nor disagree *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (Range)	Not answered *n* (%)
The new Program-level Outcome Monitoring Procedures were similar to the previous outcome monitoring procedures for the IHP.	1 (2.5%)	5 (12.5%)	10 (25%)	23 (57.5%)	2 (2.5%)	4 (1–5)	0 (0%)
The new Program-level Outcome Monitoring Procedures were an improvement over the current procedure.	2 (5%)	4 (5%)	17 (42.5%)	13 (32.5%)	5 (12.5%)	3 (1–5)	1 (2.5%)
I found the MacArthur-Bates Communicative Development Inventories Words & Gestures to be a high-quality clinical outcome evaluation tool.	1 (2.5%)	6 (15%)	12 (30%)	17 (42.5%)	1 (2.5%)	4 (1–5)	3 (7.5%)
I found the MacArthur-Bates Communicative Development Inventories Words & Gestures to be a valid and reliable tool for preschoolers with permanent hearing loss.	2 (5%)	4 (10%)	14 (35%)	15 (37.5%)	2 (5%)	4 (1–5)	3 (7.5%)
I felt the MacArthur-Bates Communicative Development Inventories Words & Gestures was the right choice for evaluating spoken language outcomes for the IHP's youngest children.	2 (5%)	3 (7.5%)	10 (25%)	17 (42.5%)	5 (12.5%)	4 (1–5)	3 (7.5%)
I found the Preschool Language Scales-5th Edition to be a high-quality clinical outcome evaluation tool.	1 (2.5%)	6 (15%)	14 (35%)	14 (35%)	3 (7.5%)	3 (1–5)	2 (5%)
I found the Preschool Language Scales-5th Edition to be a valid and reliable tool for preschoolers with permanent hearing loss.	1 (2.5%)	6 (15%)	14 (35%)	14 (35%)	3 (7.5%)	4 (1–5)	2 (5%)
I felt the Preschool Language Scales-5th Edition was the right choice for evaluating spoken language outcomes for older children in the IHP.	1 (2.5%)	8 (20%)	13 (32.5%)	14 (35%)	2 (5%)	4 (1–5)	2 (5%)
I do not have concerns about the validity/reliability of the Preschool Language Scales-5th Edition	3 (7.5%)	8 (20%)	11 (27.5%)	14 (35%)	3 (7.5%)	4 (1–5)	1 (2.5%)
I feel that implementing the new Program-level Outcome Monitoring Procedures will result in a systematic evaluation of spoken language outcomes in children with hearing loss in the IHP.	2 (5%)	1 (2.5%)	13 (32.5%)	19 (47.5%)	4 (10%)	4 (1–5)	1 (2.5%)

Note. IHP = Infant Hearing Program.

Appendix E

Speech Language Pathologists' (SLP) Opinions on the Tier 2 Procedure Statement

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neutral *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (Range)	Not applicable *n* (%)
The IVT procedures were useful for improving services for families of children with hearing loss.	2 (9.5%)	1 (4.7%)	4 (19%)	13 (61%)	0 (0%)	4 (1–4)	3 (12.5%)
The IVT procedures were useful for my clinical practice.	2 (9.5%)	2 (9.5%)	2 (9.5%)	14 (66.7%)	0 (0%)	4 (1–4)	3 (12.5%)
I was able to consistently implement the IVT procedures in my practice.	3 (13.6%)	5 (22.7%)	3 (13.6%)	10 (45.5%)	0 (0%)	4 (1–4)	2 (8.3%)

Note. IVT = Individual Vulnerability Test.

Appendix F

Speech Language Pathologists' (SLP) Perceptions of Time Involved in Tier 2 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neutral *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (Range)	Not applicable *n* (%)
The length of time it took to administer the Individual Vulnerability Tests was appropriate for incorporation into routine clinical practice.	5 (22.7%)	3 (13.6%)	4 (18.2%)	9 (40.9%)	0 (0%)	4 (1–4)	2 (8.3%)
The length of time it took to score and interpret the results of the Individual Vulnerability Tests was appropriate for incorporating into routine clinical practice.	2 (9.1%)	4 (18.2%)	3 (13.6%)	12 (54.5%)	0 (0%)	4 (1–4)	2 (8.3%)
The length of time it took to talk with parents about results of the Individual Vulnerability Tests was appropriate for incorporation into clinical practice.	2 (9.5%)	2 (9.5%)	6 (28.6%)	10 (47.6%)	0 (0%)	4 (1–4)	3 (12.5%)
The time it took to do the Individual Vulnerability Testing and reporting negatively impacted other areas of my practice.	0 (0%)	8 (36.4%)	8 (36.4%)	4 (18.2%)	1 (4.5%)	3 (2–5)	2 (8.3%)

Appendix G

Speech Language Pathologists' (SLP) Perceptions of Practice Environment for the Tier 2 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neutral *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (Range)	Not applicable *n* (%)
The environment in which I work will made it difficult for me to implement the IVT procedures.	0 (0%)	14 (63.6%)	6 (27.3%)	1 (4.5%)	0 (0%)	2 (2–4)	2 (8.3%)
I had the supplies I needed (e.g., test forms) to implement the new IVT procedures.	0 (0%)	1 (4.5%)	0 (0%)	12 (54.5%)	8 (36.4%)	4 (2–5)	2 (8.3%)
When I had questions about the IVT procedures, I consulted my colleagues.	0 (0%)	1 (5.3%)	3 (15.8%)	12 (63%)	2 (10.5%)	4 (2–5)	5 (20.8%)
When I had questions about the IVT procedures, I consulted my manager/ administrators.	5 (19%)	1 (5.3%)	7 (36.8%)	3 (15.8%)	7 (36.8%)	2 (1–4)	5 (20.8%)
When I had questions about the IVT procedures, I consulted the "Pilot Implementation Q&A" section of Western's OWL site.	1 (4.5%)	4 (18.2%)	4 (18.2%)	9 (40.9%)	3 (13.6%)	4 (1–5)	2 (8.3%)
I had the resources I needed (e.g., administrative support for scheduling, data entry) to do the IVT Procedures.	3 (21%)	1 (4.8%)	6 (28.6%)	7 (33.3%)	1 (4.8%)	4 (1–5)	3 (12.5%)
I had permission from my manager to take the time I needed to complete IVT Procedures.	0 (0%)	0 (0%)	5 (23.8%)	13 (61.9%)	2 (9.5%)	4 (3–5)	3 (12.5%)
Getting timely feedback from experts (e.g., the research team at Western University) helped me to implement the IVT Procedures.	0 (0%)	1 (5.3%)	10 (52.6%)	6 (31.6%)	1 (5.3%)	3 (2–5)	5 (20.8%)
The SLPs I work with were excited about the new IVT Procedures.	6 (27.3%)	3 (13.6%)	7 (31.8%)	4 (18.2%)	1 (4.5%)	3 (1–5)	2 (8.3%)
Managers/ administrators I work with were supportive of IVT procedures.	0 (0%)	0 (0%)	5 (23.8%)	14 (66.7%)	1 (4.7%)	4 (3–5)	3 (8.3%)

Appendix G (cont.)

Speech Language Pathologists' (SLP) Perceptions of Practice Environment for the Tier 2 Procedure

Statement	Strongly Disagree *n* (%)	Disagree *n* (%)	Neutral *n* (%)	Agree *n* (%)	Strongly Agree *n* (%)	Mode (range)	Not applicable *n* (%)
The parents I worked with were interested in the results of IVT procedures.	2 (9.1%)	3 (13.6%)	9 (41%)	6 (27.3%)	1 (4.5%)	3 (1–5)	2 (8.3%)
The task of completing the MacArthur-Bates Communicative Development Inventories - Words & Sentences was not too difficult for parents (respondents) to perform.	0 (0%)	4 (21%)	3 (15.8%)	11 (57.9%)	0 (0%)	4 (2–4)	5 (20.8%)
The task of completing the MacArthur-Bates Communicative Development Inventories - Words & Sentences was not too time consuming for parents (respondents) to perform.	0 (0%)	5 (26%)	2 (10.5%)	11 (57.9%)	0 (0%)	4 (2–4)	5 (20.8%)
The task of completing the CELF-P2 Pre-literacy Rating Scale was not too difficult for parents (respondents) to perform.	0 (0%)	0 (0%)	3 (42.8%)	2 (28.6%)	1 (14.3%)	3 (3–5)	17 (70.8%)
The task of completing the CELF-P2 Pre-literacy Rating Scale was not too time consuming for parents (respondents) to perform.	0 (0%)	1 (14.3%)	3 (42.9%)	1 (14.3%)	1 (14.3%)	3 (2–5)	17 (70.8%)

Note. IVT = Individual Vulnerability Test; CELF-P2 = Clinical Evaluation of Language Fundamentals.

Appendix H

Speech Language Pathologists' (SLP) Opinions of their Capacity to Implement the Tier 2 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neutral *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (range)	Not applicable *n* (%)
Over the past year I felt I had the clinical skills required to implement the new IVT procedures.	0 (0%)	0 (0%)	3 (13%)	12 (52%)	7 (30%)	4 (3–5)	1 (4.2%)
I am familiar with the administration of the Goldman Fristoe Test of Articulation (GFTA-3).	0 (0%)	0 (0%)	0 (0%)	9 (39%)	13 (56.5%)	5 (4–5)	1 (4.2%)
I was able to accurately score and use the norms tables for the Goldman Fristoe Test of Articulation (GFTA-3).	0 (0%)	0 (0%)	1 (5.3%)	8 (42.1%)	9 (47.4%)	5 (3–5)	5 (20.8%)
I am familiar with the administration of the MacArthur-Bates Communicative Development Inventories - Words & Sentences	0 (0%)	0 (0%)	5 (21.8%)	12 (52.2%)	5 (21.7%)	4 (3–5)	1 (4.2%)
I was able to accurately score and use the norms tables for the MacArthur-Bates Communicative Development Inventories - Words & Sentences	0 (0%)	3 (15.8%)	4 (21.1%)	8 (42%)	3 (15.8%)	4 (2–5)	5 (20.8%)
I am familiar with the administration of the Expressive One Word Picture Vocabulary Test-4th Edition (EOWPVT-4).	0 (0%)	0 (0)	4 (20%)	12 (60%)	3 (15%)	4 (3–5)	4 (16.7)
I was able to accurately score and use the norms tables for the Expressive One Word Picture Vocabulary Test-4th Edition (EOWPVT-4).	0 0%)	0 (0%)	2 (13.3%)	8 (53%)	4 (26.7%)	4 (3-5)	9 (37.5%)
I am familiar with the administration of the Clinical Evaluation of Language Fundamentals, Preschool-Second Edition (CELF-P2) Word Structure subtest.	0 (0%)	0 (0%)	0 (0%)	8 (34.8%)	14 (61.9%)	5 (4–5)	1 (4.2)
I was able to accurately score and use the norms tables for the Clinical Evaluation of Language Fundamentals, Preschool-Second Edition (CELF-P2) Word Structure subtest.	0 (0%)	0 (0%)	0 (0%)	7 (26.8%)	11 (57.9%)	5 (4–5)	5 (20.8%)

Appendix H (cont.)

Speech Language Pathologists' (SLP) Opinions of their Capacity to Implement the Tier 2 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neutral *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (range)	Not applicable *n* (%)
I am familiar with the administration of the Comprehensive Assessment of Spoken Language-Second Edition (CASL-2) Grammatical Morphemes subtest.	3 (20%)	7 (46.7%)	1 (6.7%)	3 (20%)	0 (0%)	2 (1–4)	9 (37.5%)
I was able to accurately score and use the norms tables for the Comprehensive Assessment of Spoken Language-Second Edition (CASL-2) Grammatical Morphemes subtest.	0 (0%)	1 (14.3%)	3 (42.8%)	2 (28.6%)	0 (0%)	3 (2–4)	17 (70.8%)
The new IVT procedures helped with my clinical decision-making.	2 (9.5%)	2 (9.5%)	4 (19.05%)	9 (42.9%)	3 (14.3%)	4 (1–5)	3 (12.5%)
The new IVT procedures helped parents with their decision-making.	2 (9.5%)	3 (14.3%)	8 (38.1%)	7 (33%)	0 (0%)	3 (1–4)	3 (12.5%)
Repeat administration of the Individual Vulnerability tests to the same child 6–12 months later benefited the families and children that I serve.	1 (5.3%)	4 (21%)	4 (21%)	7 (37%)	2 (10.5%)	4 (1–5)	5 (20.8%)

Note. IVT = Individual Vulnerability Test.

Appendix I

Speech Language Pathologists' (SLP) Perceptions of the Quality of the Tier 2 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neutral *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (range)	Not applicable *n* (%)
I found the assessment tools required for the IVT to be high quality clinical outcome evaluation tools.	0 (0%)	1 (4.5%)	3 (13.6%)	14 (64.6%)	3 (13.6%)	4 (2–5)	2 (8.3%)
I felt the MacArthur-Bates CDI Words and Gestures "Words Produced" was the right choice for evaluating vocabulary vulnerability in children with permanent hearing loss (8–18 months).	2 (10%)	2 (10%)	4 (20%)	8 (40%)	3 (15%)	4 (1–5)	4 (16.7%)
I felt the MacArthur-Bates CDI Words and Sentences "Words Produced" was the right choice for evaluating vocabulary vulnerability in children with permanent hearing loss (19–30 months).	3 (15%)	2 (10%)	4 (20%)	7 (35%)	3 (15%)	4 (1–5)	4 (16.7%)
I felt the Expressive One Word Picture Vocabulary Test (EOWPVT-4) was the right choice for evaluating vocabulary vulnerability in children with permanent hearing loss (24–35 months).	0 (0%)	1 (6.3%)	6 (37.5%)	7 (43.8%)	1 (6.3%)	4 (2–5)	8 (33.3%)
I felt the CELF-P2 Word Structure subtest was the right choice for evaluating grammar vulnerability in children with permanent hearing loss (3–6 years).	0 (0%)	1 (5.3%)	4 (21%)	10 (52.6%)	3 (15.8%)	4 (2–5)	5 (20.8%)
I felt the CASL-2 Grammatical Morphemes subtest was the right choice for evaluating grammar vulnerability in children with permanent hearing loss (3–6 years).	0 (0%)	0 (0%)	4 (66.7%)	1 (16.7%)	0 (0%)	3 (3–4)	18 (75%)
I felt the Goldman Fristoe Test of Articulation, Third Edition (GFTA-3) - Sounds in Words subtest was the right choice for evaluating vocabulary and syntax vulnerability in children with permanent hearing loss (30–48 months).	1 (5.6%)	3 (16.7%)	3 (16.7%)	7 (38.9%)	3 (16.7%)	4 (1–5)	6 (25%)

Appendix I (cont.)

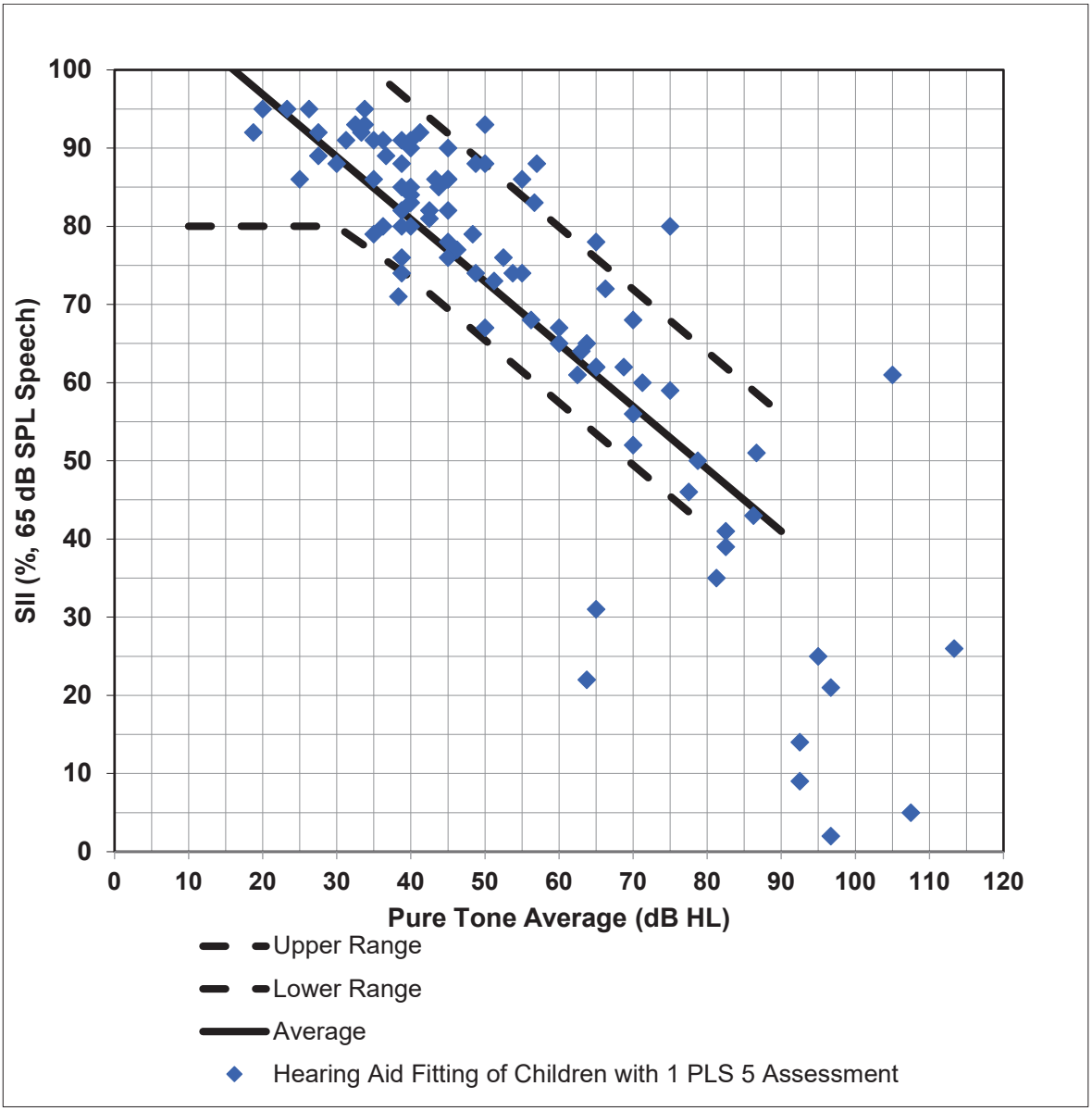
Speech Language Pathologists' (SLP) Perceptions of the Quality of the Tier 2 Procedure

Statement	Strongly disagree *n* (%)	Disagree *n* (%)	Neutral *n* (%)	Agree *n* (%)	Strongly agree *n* (%)	Mode (range)	Not applicable *n* (%)
I felt the CELF-P2 Pre-literacy rating scale was the right choice for evaluating emergent literacy/ phonological awareness vulnerability in children with permanent hearing loss (4–6 years).	0 (0%)	1 (7.7%)	6 (46%)	5 (38.5%)	0 (0%)	3 (2–4)	11 (45.8%)
I felt the CELF-P2 Phonological Awareness subtest was the right choice for evaluating emergent literacy/phonological awareness vulnerability in children with permanent hearing loss (4–6 years).	0 (0%)	0 (0%)	6 (37.5%)	8 (50%)	1 (6.25%)	4 (3–5)	8 (33%)
I feel the implementation of IVT helped me to identify impairments in children with permanent hearing loss that were missed through Program Level Outcome Monitoring.	4 (20%)	2 (10%)	1 (5%)	10 (50%)	2 (10%)	4 (1–5)	4 (16.7%)

Note. IVT = Individual Vulnerability Test.

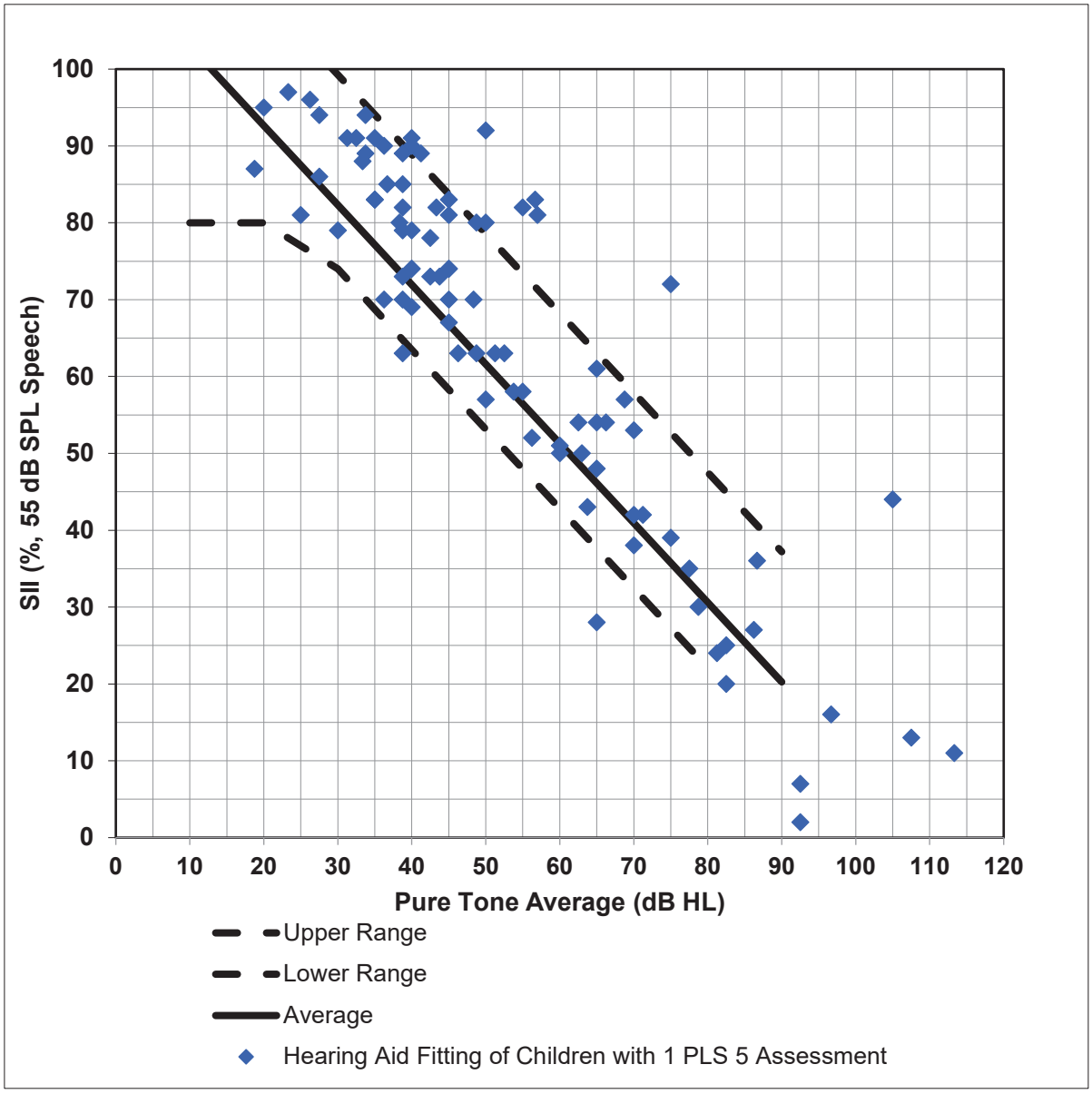
Supplemental Material 1

Hearing Aid Speech Intelligibility Index (SII) at 65 dB compared to Moodie et al., 2017 normative data: Children with data from the Preschool Speech & Language Scale (PLS), 5th ed



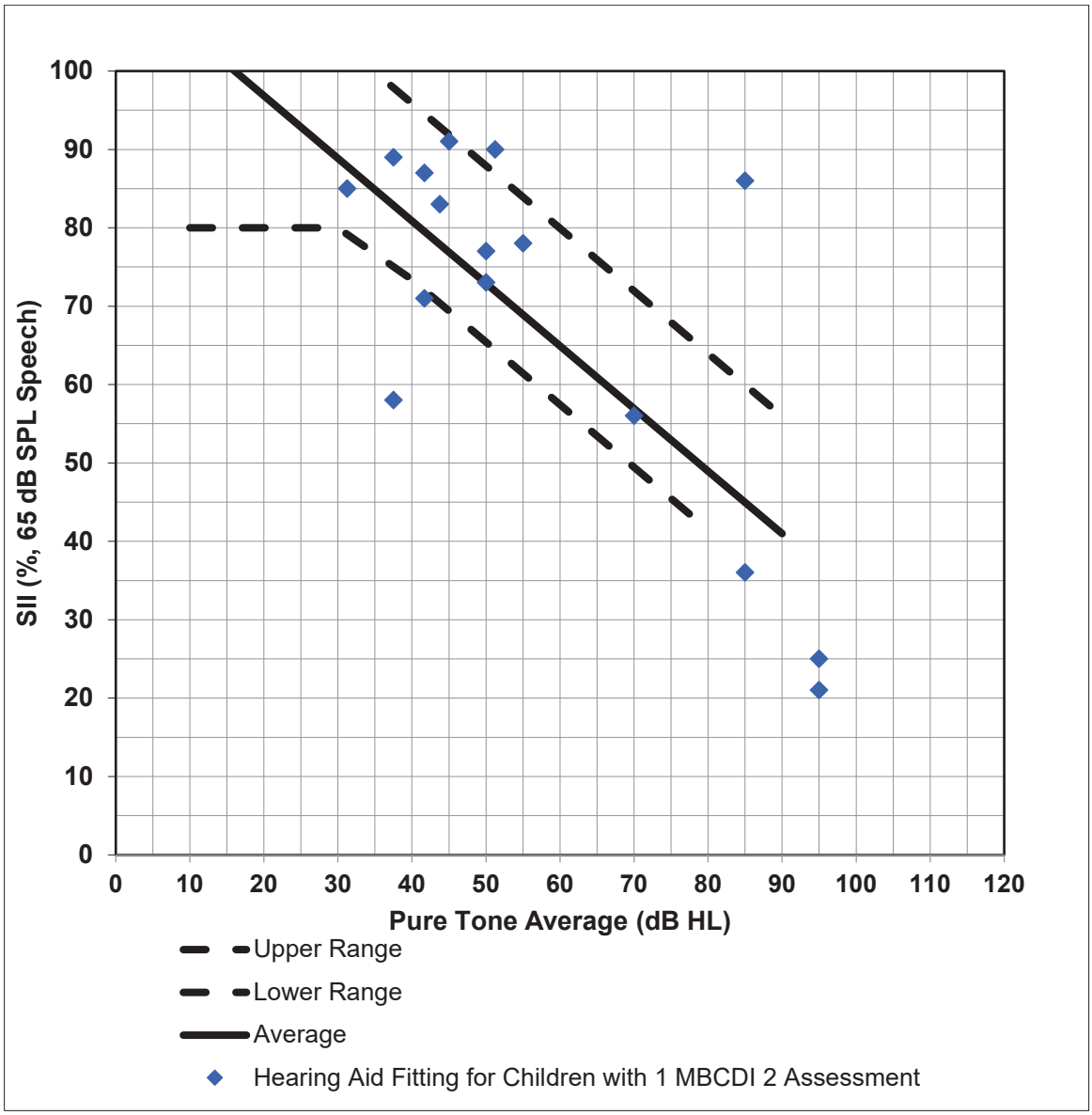
Supplemental Material 2

Hearing Aid Speech Intelligibility Index (SII) at 55 dB compared to Moodie et al., 2017 normative data: Children with data from the Preschool Speech & Language Scale (PLS), 5th ed



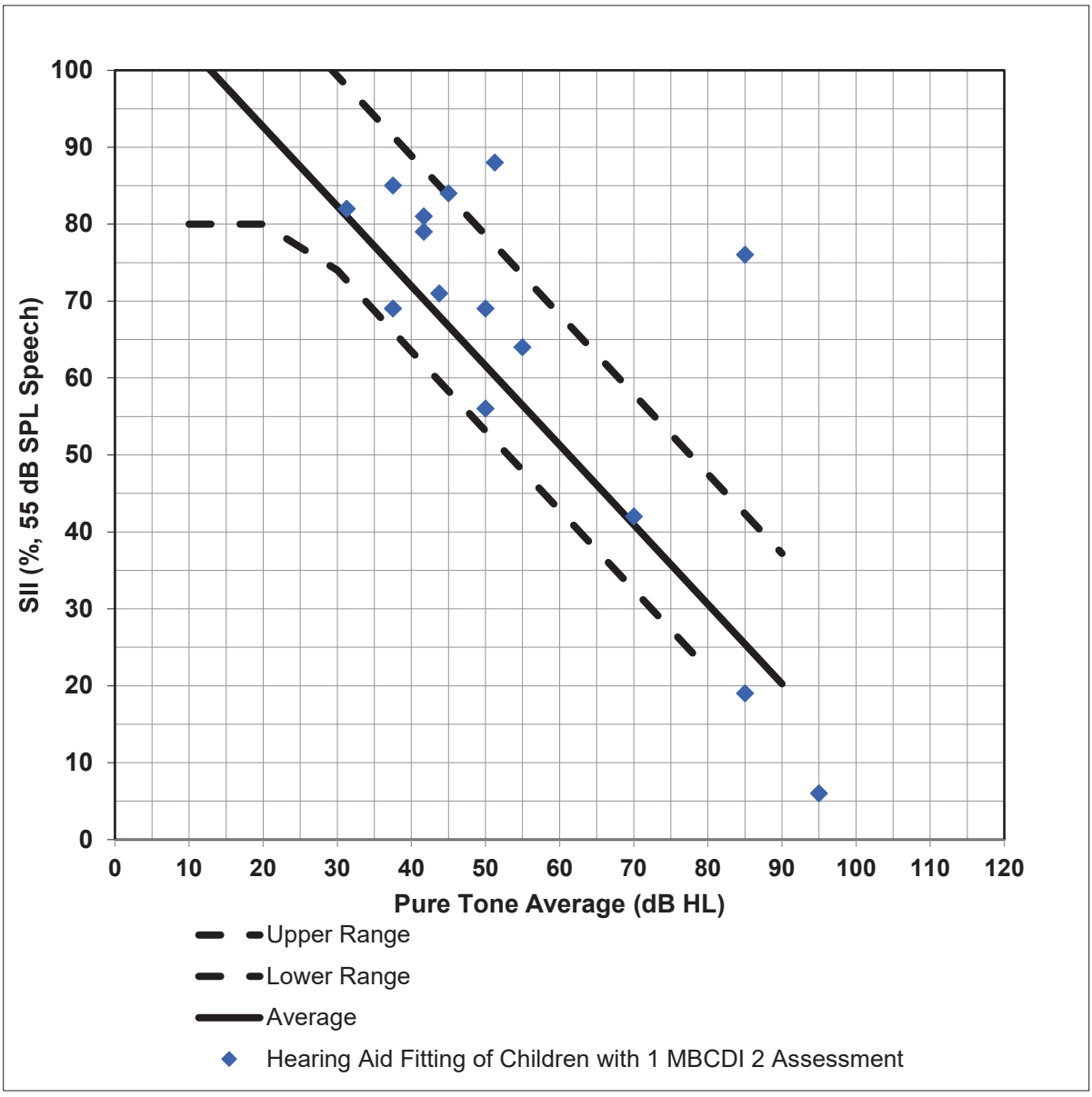
Supplemental Material 3

Hearing Aid Speech Intelligibility Index (SII) at 65 dB compared to Moodie et al., 2017 normative data: Children with data from the MacArthur-Bates Communicative Development Inventories, 2nd ed. (MBCDI-2)



Supplemental Material 4

Hearing Aid Speech Intelligibility Index (SII) at 65 dB compared to Moodie et al., 2017 normative data: Children with data from the MacArthur-Bates Communicative Development Inventories, 2nd ed. (MBCDI-2)



Supplemental Material 5

Distributions of Preschool Language Scale, 5th ed. (PLS-5) Scores