Developing a Spoken Language Outcome Monitoring Procedure for Early Hearing Detection and Intervention Programs

Olivia Daub, The University of Western Ontario

Supervisor: Oram Cardy, Janis, The University of Western Ontario
Co-Supervisor: Bagatto, Marlene P, The University of Western Ontario

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Health and Rehabilitation Sciences

© Olivia Daub 2021

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Recommended Citation
https://ir.lib.uwo.ca/etd/7917

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.
Abstract

Early Hearing Detection and Intervention programs are associated with improved spoken language outcomes for children who are deaf/hard-of-hearing. Best practice recommendations call for regular spoken language outcome monitoring to support decision making for all stakeholders (families, audiologists, speech-language pathologists, and program managers).

Despite the clear calls for spoken language outcome monitoring, there is no peer-reviewed guidance as to how Early Hearing Detection and Intervention programs can best accomplish this monitoring. This dissertation evaluates the assumptions underlying spoken language outcome monitoring and contributes a new procedure developed for a Canadian Early Hearing Detection and Intervention program: the Ontario Infant Hearing Program.

Whether decisions can be validly made using assessment data underpins the tenability of spoken language outcome monitoring. Chapter 2 considers test misuse across the profession of speech-language pathology from test design to clinical practice. I argue that a conceptual validity framework is one potential solution. This framework is applied throughout the dissertation.

Chapter 3 aims to develop a spoken language outcome monitoring procedure to support the Ontario Infant Hearing Program. This chapter describes the process I engaged in, including a scoping review and critical appraisal of norm-referenced spoken language tests, to develop an outcome monitoring procedure for the Infant Hearing Program.

Prior to implementing the recommended procedures province-wide, the Infant Hearing Program needed evidence as to whether the recommendations (a) meaningfully inform stakeholder decisions and (b) are feasible to implement. Chapter 4 reports on a pilot implementation of the recommended procedures and speech-language pathologists’ perceptions of it.

During development of the procedure outlined in Chapter 3, one of the key vulnerabilities I recommended to monitor was early vocal development in children who are younger than 2 years. Chapter 5 is a survey study capturing the clinical questions speech-language pathologists’ have about early vocal development of children who are deaf/hard-of-hearing to inform future projects to assess the validity of candidate vocal development assessments.

Overall, this dissertation contributes a spoken language outcome monitoring procedure for Early Hearing Detection and Intervention programs and highlights the tension between decisions, psychometrics, and implementation, in accomplishing spoken language outcome monitoring to inform best practice recommendations.
Keywords

Speech language pathology; Early Hearing Detection and Intervention Programs;
Permanent Childhood Hearing Loss; Spoken Language Outcome Monitoring; Assessment;
Psychometrics; Test Validation; Implementation Science; Knowledge Translation
Summary for Lay Audience

Early Hearing Detection and Intervention programs are designed to help children who are deaf/hard-of-hearing access language (spoken or signed) early in development so that they do not have difficulty learning language when they are older. For children who are learning spoken languages, best practice recommendations say that these programs should regularly measure language development so that they know that services are effective, and children are on track for learning spoken language. However, there is currently no research that describes how programs should measure spoken language, whether the tools that speech-language pathologists need exist, and whether the tools are easily used in practice.

This dissertation aimed to identify whether it is currently possible to measure spoken language development in children with hearing loss and design a method to measure spoken language development for a Canadian Early Hearing Detection and Intervention program. This dissertation explores how stakeholders (i.e., speech-language pathologists and government managers) use tests to make decisions and applies this framework to the development of a new spoken language outcome monitoring procedure. Then, this dissertation evaluates whether the new procedure results in data that are usable for program evaluation and suitable to implement in clinical practice. Finally, this dissertation documents the questions that speech-language pathologists have about children who are deaf/hard-of-hearing’s vocal development to inform the design of new approaches to incorporate into the outcome monitoring procedure.
Overall, this dissertation highlights the complexities of achieving spoken language outcome monitoring and recommendations for Early Hearing Detection and Intervention programs looking to develop their own procedures.
List of Abbreviations

BEPTA: Better-ear Pure Tone Average

CASL: Comprehensive Assessment of Spoken Language

CDHH: Children who are deaf/hard-of-hearing

CDI: Child Development Inventory

CELF: Comprehensive Evaluation of Language Fundamentals

CHH: Children who are hard-of-hearing

COSMIN: Consensus Based Standards for the Selection of Health Status Measurement Instruments

DEAP: Diagnostic Evaluation of Articulation and Phonology

DHH: deaf/hard-of-hearing

EHDI: Early Hearing Detection and Intervention

EOWPVT: Expressive One Word Vocabulary Test

EVT: Expressive Vocabulary Test

GFTA: Goldman-Fristoe Test of Articulation

IHP: Infant Hearing Program
IMP: Infant Monitor of Vocal Productions

JCIH: Joint Committee on Infant Hearing

KLPA: Khan-Lewis Phonological Analysis

LEESPDQ: LittlEARs Early Speech Production Questionnaire

MBCDI: MacArthur Bates Communicative Development Inventories

(M)CDI: (Minnesota) Child Development Inventory

MSEL: Mullen Scales of Early Learning

OMRU: Ottawa Model of Research Use

PLAI: Preschool Language Assessment Inventory


PPVT: Peabody Picture Vocabulary Test

PRISE: Prelexical Infant Scale Evaluation

SII: Speech Intelligibility Index

SLP: Speech-language pathologist

TACL: Test of Auditory Comprehension of Language

UNHS: Universal Newborn Hearing Screening
VABS: Vineland Adaptive Behavior Scales

VDLI: Vocal Development Landmarks Interview
Co-Authorship Statement

Chapter 2: 

Chapter 3: 

Chapter 4: 

Chapter 5: 

Chapters 2, 3, 4, and 5 included in this dissertation are manuscripts led by me (Olivia Daub; the first listed author) and co-authored by my collaborators. In all chapters, I led the data collection, analysis, manuscript preparation, and submission. My collaborators contributed to the development of the partnerships involved in these projects, securing funding, methodological support, and contributing to the final drafts of the manuscripts. I am the sole author of Chapters 1 and 6 and received support from my supervisors Drs. Janis Oram Cardy and Marlene Bagatto on final drafts.
Acknowledgments

As my graduate training comes to a close, I have found myself increasingly reflecting on the importance of mentorship within academia. I have been remarkably privileged during my training to have received support from a broad network of incredible mentors who have invested themselves significantly in my development as a clinician and scientist. First and foremost, thank you to my supervisors Drs. Janis Oram Cardy and Marlene Bagatto. You both have been fierce proponents of my career since I joined Western 7 years ago and have always encouraged me to tackle the research problems that I am most passionate about. The past 7 years have flown by.

Thank you to my advisory committee, Drs. Lisa Archibald and Susan Scollie, who challenged me every step of the way. This dissertation is richer for your contributions.

To Dr. Katherine White, my first mentor – thank you. As you know, I flatly rejected the idea of a research career prior to joining your lab. None of this work would have happened without your gentle encouragement to continue asking questions and patiently pursuing answers.

Along the way, I have been very fortunate to be supported by members of Western’s faculty. Dr. BJ Cunningham has been an amazing collaborator and colleague, who has patiently entertained my quest for side projects and new research partnerships. Dr. Andrew Johnson has shaped my academic career immeasurably, having piqued my interest in statistics and psychometrics early in my training. Dr. Sheila Moodie, thank you for our conversations and your enthusiastic support. Dr. Betsy Skarakis-Doyle, thank you for our wonderful coffee chats and challenging me on my assumptions underlying language development and measurement.

To the Ontario Ministry of Children, Community and Social Services – thank you for allowing me to be a part of this process. Your partnership has been invaluable. Thank you also to the amazing speech language pathologists who supported these projects.

Thank you to the members of the ASLD and shared lab: Elaine, Rachael, Caitlin, Meghan, Alex, Taylor, Alyssa, and Theresa for your support through both clinical and research stages of our training. Your friendship and our shared lab discussions have been particularly missed during COVID-19. Thank you also to Noa, Nancy, Jess D. Alex, Jess K., and Regan for your friendship and support.

To my friends and family who have shown me love and support through my graduate training: thank you. To my dad, who consistently pushed me to aim high and work hard, thank you. Melodie and John, surviving the last year of my training without your support and help with Imogen would not have been possible. And finally, to Oliver – who may, at this point, know my dissertation better than I do. Your love and support have kept me grounded and focused. “Thank you” will never be enough.
Dedication

This dissertation is dedicated to the two most important women in my life: my mother, Penny Daub (1962-2010), who started me on my academic journey and my daughter, Imogen Bailey (2020-), who arrived in time to see its conclusion. My mom’s greatest gift to me was her faith that I can achieve whatever it is I set my mind to. Imogen, it is my hope that I can do the same for you.
# Table of Contents

Abstract ........................................................................................................................................... ii

Summary for Lay Audience ................................................................................................................. iv

List of Abbreviations ............................................................................................................................ vi

Co-Authorship Statement ....................................................................................................................... ix

Acknowledgments ................................................................................................................................. x

Dedication ........................................................................................................................................... xi

Table of Contents ................................................................................................................................. xii

List of Tables ...................................................................................................................................... xvii

List of Figures ..................................................................................................................................... xix

List of Appendices ............................................................................................................................... xx

Chapter 1 ............................................................................................................................................. 1

1 Introduction ....................................................................................................................................... 1

1.1 Permanent Childhood Hearing Loss & Spoken Language Development ...................... 2

1.2 Early Hearing Detection and Intervention Programs: History and Evidence of Effectiveness ................................................................. 4

1.3 Early Hearing Detection and Intervention Programs: Best Practice Recommendations ..................................................................................... 5

1.4 Spoken Language Outcome Monitoring: Psychometric Considerations .................. 7

1.5 Spoken Language Outcome Monitoring: Implementation Considerations .......... 11

1.6 Spoken Language Outcome Monitoring in The Ontario Infant Hearing Program ..... 13

1.7 The Present Studies: Purpose & Motivation ........................................................................... 15

References ........................................................................................................................................ 18

Chapter 2 ........................................................................................................................................... 21
2 Adopting a Conceptual Validity Framework for Testing in Speech-Language Pathology ................................................................. 21

2.1 Underlying Concepts of the Validity Framework .................................................. 25

2.2 The Validity Framework .................................................................................. 27

2.2.1 Validity Framework Step 1: Identify the Clinical Decisions ................. 27

2.2.2 Validity Framework Step 2: Identifying What Evidence is Needed to Support the Decision ......................................................... 29

2.2.3 Validity Framework Step 3: Evaluating the Evidence ................................. 30

2.2.4 Integrating the Validity Framework in Practice ........................................ 38

2.3 Example Decisions and Necessary Validity Evidence .................................. 40

2.4 Benefits of Adopting the Standard’s framework ........................................... 44

2.5 Next Steps Towards Adopting the Validity Framework ................................ 48

2.6 Conclusions .................................................................................................. 54

References ........................................................................................................ 55

Chapter 3 ............................................................................................................ 58

3 Developing a Spoken Language Outcome Monitoring Procedure for a Canadian Early Hearing Detection and Intervention Program: Process and Recommendations .......... 58

3.1 Step 1: Identifying Assessment Purposes ...................................................... 63

3.1.1 The IHP’s Assessment Purposes at the Program Level .......................... 63

3.1.2 The IHP’s Assessment Purposes for Individual Children and Families .. 65

3.1.3 Matching the Assessment Purpose with the Assessment Method .......... 66

3.2 Step 2: Selecting Tests for Outcome Monitoring ......................................... 67

3.2.1 Step 2a) Scoping Review of Norm-Referenced Tests ............................. 67

3.2.2 Step 2b) Critical Appraisal of Norm-Referenced Tests ......................... 74

3.2.3 Step 2c) Consideration of Administration Properties ............................ 77
3.3 Step 3: Integrating the Evidence into Recommendations ........................................ 79

3.3.1 Recommendations for Overall Spoken Language Outcome Monitoring .................. 79
3.3.2 Recommendations for Individual Vulnerability Testing ...................................... 81

3.4 Consultation with Stakeholders .............................................................................. 82
3.5 Discussion ............................................................................................................. 83
3.6 Conclusions .......................................................................................................... 91

References .................................................................................................................. 93

Chapter 4 ..................................................................................................................... 99

4 Usability and feasibility of a Spoken Language Outcome Monitoring Procedure in a
Canadian Early Hearing Detection & Intervention Program: Results of a 12-month
Pilot ............................................................................................................................. 99

4.1 Pilot Study 1: Tier 1 Program-level Outcome Monitoring .................................... 105

4.1.1 Method .............................................................................................................. 105
4.1.2 Analyses: Data usability ................................................................................. 111
4.1.3 Analyses: Procedure feasibility ................................................................. 112
4.1.4 Results: Data usability ............................................................................... 113
4.1.5 Results: Procedure feasibility ................................................................. 115

4.2 Pilot Study 2 – Tier 2 Individual Vulnerability Testing ....................................... 115

4.2.1 Method .......................................................................................................... 115
4.2.2 Analysis: Data usability ............................................................................. 116
4.2.3 Analysis: Procedure feasibility ................................................................. 118
4.2.4 Results: Data usability ............................................................................... 118
4.2.5 Results: Procedure feasibility ................................................................. 119

4.3 Discussion ........................................................................................................... 120

4.3.1 Data Usability ............................................................................................... 121
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.2</td>
<td>Procedure Feasibility</td>
<td>125</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Limitations &amp; Future Directions</td>
<td>126</td>
</tr>
<tr>
<td>4.4</td>
<td>Conclusions</td>
<td>128</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>129</td>
</tr>
<tr>
<td>Chapter 5</td>
<td></td>
<td>132</td>
</tr>
<tr>
<td>5</td>
<td>What do Speech-Language Pathologists want to know when assessing early vocal development in children who are (D)eaf/hard-of-hearing?</td>
<td>132</td>
</tr>
<tr>
<td>5.1</td>
<td>Existing vocal development tests</td>
<td>134</td>
</tr>
<tr>
<td>5.1.1</td>
<td>LittLEARs Early Speech Production Questionnaire® (LEESPQ®)</td>
<td>134</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Vocal development Landmarks Interview (VDLI)</td>
<td>135</td>
</tr>
<tr>
<td>5.1.3</td>
<td>Infant Monitor of Vocal Productions (IMP)</td>
<td>136</td>
</tr>
<tr>
<td>5.1.4</td>
<td>Prelexical Infant Scale Evaluation (PRISE)</td>
<td>136</td>
</tr>
<tr>
<td>5.2</td>
<td>What Validity Evidence is Needed?</td>
<td>137</td>
</tr>
<tr>
<td>5.3</td>
<td>Study Purpose</td>
<td>139</td>
</tr>
<tr>
<td>5.4</td>
<td>Methods</td>
<td>141</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Ethical Approval</td>
<td>141</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Participants</td>
<td>141</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Online Survey</td>
<td>143</td>
</tr>
<tr>
<td>5.5</td>
<td>Analysis</td>
<td>146</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Vocal Development Assessment Purposes</td>
<td>146</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Barriers to Vocal Development Assessment</td>
<td>147</td>
</tr>
<tr>
<td>5.6</td>
<td>Results</td>
<td>147</td>
</tr>
<tr>
<td>5.6.1</td>
<td>Assessment Purposes</td>
<td>147</td>
</tr>
<tr>
<td>5.6.2</td>
<td>Barriers to Vocal Assessment</td>
<td>150</td>
</tr>
</tbody>
</table>
List of Tables

Table 2-1 Considerations for determining the presence/absence of disorder ...... Error! Bookmark not defined.

Table 2-2 Considerations for determining change in skills ....... Error! Bookmark not defined.

Table 2-3 Considerations for determining severity/level of function ............... Error! Bookmark not defined.

Table 3-1 Norm-referenced test use in research with children who are hard of hearing ................................................................. 73

Table 3-2 Critical appraisal of norm-referenced tests using COSMIN criteria ........ 77

Table 3-3 Administration properties for currently available versions of psychometrically suitable for norm-referenced tests ................................................................. 79

Table 4-1 Demographics of children with data for one PLS-5 assessment .......... 109

Table 4-2 Demographics of children with data for two PLS-5 assessments .......... 109

Table 4-3 Demographics of children with data for one MBCDI assessment .......... 110

Table 4-4 Demographics of children with data for two MBCDI assessments .......... 110

Table 4-5 Relation of PLS-5 standard scores with predictors .......................... 113

Table 4-6 Relation of MBCDI-2 percentile ranks with predictors ...................... 114
Table 4-7 Demographics of children included in pilot study 2 ................................. 117

Table 4-8 Agreement between PLS-5 Auditory Comprehension and GFTA-3 Sounds-in-Words Subtest ................................................................. 118

Table 4-9 Agreement between PLS-5 Expressive Communication and GFTA-3 Sounds-in-Words Subtest ................................................................. 119

Table 4-10 Agreement between PLS-5 Auditory Comprehension and CELF-P2 Words Structure Subtest ................................................................. 119

Table 4-11 Agreement between PLS-5 Expressive Communication and CELF-P2 Words Structure Subtest ................................................................. 119
List of Figures

Figure 2-1: The conceptual validity framework .......................................................... 24
Figure 2-2: Diagnosing disorders decision tree ......................................................... 31
Figure 2-3: Measuring progress decision tree ............................................................ 32
Figure 2-4: Determining severity decision tree ......................................................... 33
Figure 3-1: Process for developing proposed outcome monitoring process .......... 64
Figure 3-2: Proposed outcome monitoring process ................................................... 68
Figure 3-3: Articles included for evaluation ............................................................... 72
Figure 3-4: Final recommendation ............................................................................. 83
Figure 4-1: Overall outcome monitoring process, Daub & Oram Cardy (2021) ...... 102
Figure 4-2: Tests used in outcome monitoring process, Daub & Oram Cardy (2021) .................................................................................................................................................. 103
List of Appendices

Appendix 1: CINAHL Search Strategy ................................................................. 190

Appendix 2: Data extraction ............................................................................. 190

Appendix 3: Children with one PLS-5 assessment hearing aid Speech Intelligibility
Index (SII) in conversational speech, compared to Moodie et al., 2017 normative data...... 191

Appendix 4: Children with one PLS-5 assessment hearing aid Speech Intelligibility
Index (SII) in quiet speech, compared to Moodie et al., 2017 normative data................. 192

Appendix 5: Children with one MBCDI-2 assessment hearing aid Speech
Intelligibility Index (SII) in conversational speech, compared to Moodie et al., 2017
normative data........................................................................................................ 193

Appendix 6: Children with one MBCDI-2 assessment hearing aid Speech
Intelligibility Index (SII) in quiet speech, compared to Moodie et al., 2017 normative data194

Appendix 7: Tier 1 post-pilot survey ................................................................... 195

Appendix 8: SLP’s opinions on the Program-level Outcome Monitoring Procedure
Statement................................................................................................................ 209

Appendix 9: SLPs opinions of their capacity to implement the Program-level
Outcome Monitoring Procedure .............................................................................. 209
Appendix 10: SLPs’ Opinions on the practice environment and the Program-level Outcome Monitoring Procedure ................................................................. 210

Appendix 11: SLPs’ Opinions on the quality of the Program-level Outcome Monitoring Procedure ........................................................................................................... 211

Appendix 12: SLP’s opinions on the Program-level Outcome Monitoring Procedure Statement........................................................................................................ 212

Appendix 13: Tier 2 post-pilot survey ................................................................................................................................. 212

Appendix 14: SLPs’ perceptions of time involved in IVT procedure .......................................................... 222

Appendix 15: SLPs’ perceptions of practice environment for IVT procedure ........................................ 222

Appendix 16: SLPs’ opinions of their capacity to implement the IVT procedure .................................... 223

Appendix 17: SLPs’ perceptions of the quality of the IVT procedure ......................................................... 224

Appendix 18: Distribution of PLS-5 standard scores ......................................................................................... 225

Appendix 19: The Revised Ottawa Model of Research Use ............................................................................. 226

Appendix 20: Vocal development barriers and assessment purposes survey ............................................. 227
Chapter 1

1 Introduction

Children who are deaf/hard-of-hearing (DHH) who are acquiring a spoken language are faced with the challenge of learning language in the context of a sensory impairment that limits their formative experiences with the language input. Early Hearing Detection and Intervention (EHDI) programs have been developed to ensure all children who are DHH have access to language (signed or spoken) early in development to ensure they can reach their full potential. Because permanent childhood hearing loss is not a language learning disorder, when children who are DHH are able to access spoken language, families and service providers can expect that children will develop spoken language commensurate with their same-aged peers. Whether a child achieves appropriate spoken language can be used as a measure of whether the EHDI program successfully supported spoken language development. Regular assessment of a child’s spoken language development serves the dual purposes of identifying whether EHDI programs are effectively supporting children who are DHH as well as enabling service providers to identify when a child is not progressing towards age-appropriate outcomes so that they can better support the child (Joint Committee on Infant Hearing, 2007, 2013, 2019). This dissertation had two primary aims. First, it aimed to provide practical support to a Canadian EHDI program to develop and evaluate a spoken language outcome monitoring procedure prior to implementation within the program. Second, I sought to evaluate the assumptions underlying recommendations for spoken language outcome monitoring to inform best practice considerations moving forward. The present chapter provides an overview of EHDI programs and their aims, as well as the evidence available
to EHDI programs to guide the development of spoken language outcome monitoring procedures.

1.1 Permanent Childhood Hearing Loss & Spoken Language Development

Permanent childhood hearing loss is documented to impact between 1 – 3 children per 1000 live births (Mehl & Thompson, 1997). The majority of children who are DHH are born to families where one, or both, parents are hearing (Mitchell & Krachmer, 2004) and where spoken languages are used in the home.

Basic psychological research has routinely documented that early auditory and language experiences are fundamental in the spoken language development of infants and young children, even prior to birth. Access to low frequency, prosodic information in utero is speculated to support the acquisition of prosodic learning which, in turn, can facilitate later language learning tasks such as phrase and word segmentation (Moon et al., 2012). In the case of permanent childhood hearing loss, children who are DHH and learning a spoken language are faced with the immense task of learning their language in the presence of a sensory impairment that inherently limits their early auditory and spoken language experiences. As a result, permanent childhood hearing loss has been routinely associated with poor speech and language outcomes in early childhood, with cascading influences on academic and psycho-social outcomes (Moeller, 2000; Patel & Feldman, 2011).

Explanations for language learning difficulties in children who are DHH are multifactorial. First, they do not necessarily have a language learning disorder per se (although hearing loss does not preclude impairments in the language learning
mechanism). The inconsistent access hypothesis, proposed by Tomblin and Moeller (2015), argues that the interaction between the child’s own cognitive abilities, their linguistic environment, and their access to auditory information (conceptualized as an interaction between their hearing sensitivity, audibility provided by amplification, and hours per day of amplification use) shape eventual spoken language outcomes of children who are DHH. That is, the sensory limitations of a child who is DHH impose inconsistency on their access to auditory information, including access to different sounds associated with specific phonemes (most frequently high frequency sounds such as fricatives). Limits in auditory access are influenced by amplification technology-and properties of the auditory/linguistic environment (e.g., caregiver input).

Notably, the inconsistent access hypothesis highlights that inconsistent access to auditory information is ongoing and can have specific implications for certain domains of spoken language. It has been consistently documented that children who are DHH, when provided with adequate intervention (described later), can achieve overall spoken language abilities commensurate with their same-aged peers. However, even in cases when a child might score within normal limits on an omnibus measure of language development, they continue to be at risk for impairment in key domains, namely: (a) vocal development and first words, (b) articulation, (c) morphosyntax, and (d) phonemic awareness and early literacy (Moeller et al., 2007). Typical spoken language development, therefore, cannot be inferred by typical acquisition of earlier milestones because children who are DHH are faced with the ongoing task of acquiring new linguistic information in the face of inconsistent auditory information.
1.2 Early Hearing Detection and Intervention Programs: History and Evidence of Effectiveness

The development of and advocacy for EHDI programs was prompted by both the recognition of the importance of early auditory, and linguistic, experience in shaping spoken language development and the development of new technologies sensitive to support accurate screening and identification in the first year of life. The Joint Committee on Infant Hearing (JCIH) – developed in 1969 – was formed to synthesize the research evidence surrounding issues related to permanent childhood hearing loss, such as the development of hearing screening technologies, and to develop position papers to guide best practice. Over the first three decades, the JCIH released regular guidance on the use of high-risk registries and, later, universal newborn hearing screening (UNHS) in order to facilitate early identification and intervention for children with permanent childhood hearing loss.

In the 2000s, with clear evidence that early auditory and spoken language experiences matter in spoken language development and with the availability of technology to screen, identify, and improve auditory access in permanent childhood hearing loss, the JCIH developed the first recommendation for EHDI programs (JCIH, 2000). EHDI programs are family-centered systems approaches to intervention. Comprehensive EHDI programs include UNHS to support early identification, as well as comprehensive intervention services (e.g., amplification technology, speech-language pathology services, signed language services) in the family’s preferred language modality (signed or spoken language). Children who are DHH and are born in regions with EHDI programs have improved spoken language outcomes over those who are born in regions without EHDI services (Ching et al., 2013). Spoken language outcomes for children who
are fitted with hearing aids are influenced by numerous factors including the age a child is fitted, the quality of the fit, and the consistency with which children who are DHH use their hearing aids (Moeller & Tomblin, 2015; Tomblin et al., 2015). In addition, there are other factors known to impact language development such as the presence of additional disabilities (Cuppes et al., 2014), the quantity and quality of parental linguistic input (Ambrose et al., 2014), and earlier language abilities (Daub et al., 2017). Importantly, when children who are DHH (as a group) receive adequate EHDI services, they are documented to achieve spoken language outcomes that are within age expectations (e.g., Ching et al., 2013; Tomblin et al., 2015).

1.3 Early Hearing Detection and Intervention Programs: Best Practice Recommendations

As researchers continue to identify components of EHDI programming that result in improved outcomes for children who are DHH, best practice recommendations continue to be developed and refined to guide EHDI program development and provision. As new evidence surrounding the effectiveness of EHDI programs is reported, and advances in screening, assessment, and intervention technology are made, the JCIH continues to provide updated practice recommendations. These recommendations cover the breadth of services included within an EHDI program, including recommendations for UNHS, supporting families in selecting language modalities, audiological amplification/implantation (if appropriate), and service provision. These recommendations have been adopted internationally (e.g., the Canadian Infant Hearing Task Force) and the principles re-iterated and expanded on by international consensus groups (Moeller et al., 2013).
Of particular interest to this dissertation is best practice recommendations surrounding spoken language outcome monitoring as a key component of EHDI programs. Because the spoken language impairments associated with permanent childhood hearing loss result from inconsistent access to spoken language, not an impairment in the language learning mechanism, it has been argued that children who are DHH should be able to achieve language (spoken or signed) outcomes commensurate with their same aged peers (JCIH, 2013) when they access appropriate EHDI services. Some research has documented that, on some spoken language measures, children who are DHH can (as a group) perform within -1 SD of the mean of their peers (Daub et al., 2017; Fulcher et al., 2012; Moeller 2000; Stika et. al, 2014). However, even in cases where children who are DHH meet major language milestones early on in their development, inconsistent access to auditory information (influenced by factors such as hearing device usage and audibility; Moeller & Tomblin, 2015) has ongoing implications for the learning of some speech/language domains (e.g., articulation, some aspects of morphology, phonological awareness; see Moeller et al., 2007 for a review). Based on these data, international recommendations for EHDI best practices advocate for the routine monitoring of spoken language development from birth to program discharge (JCIH, 2013; Moeller et al., 2013) to: (a) determine if a child is progressing towards their goals; (b) determine if a child is developing appropriately compared to their peers; and (c) inform therapy planning. The spoken language development of children who are DHH is recommended to be assessed every 6 months until the child’s third birthday and annually thereafter. Results of these assessments are to be used to identify when a child is making appropriate progress towards developmentally appropriate language goals
(defined as within 1 SD of the mean for normal hearing peers; JCIH, 2013) and to inform clinical decision making.

These recommendations, while appropriate to ensure adequate service provision, rest on a complex series of assumptions relating to spoken language assessment, the tenability of which have not been previously evaluated. In order to implement spoken language progress monitoring, there must exist the appropriate tools to do so. In order for spoken language outcome progress monitoring to be successful in the ways conceptualized by the JCIH, the field requires tools that consider psychometric complexities and the complexity of implementing new procedures into clinical practice in order to design a procedure that balances the tensions between the two.

1.4 Spoken Language Outcome Monitoring: Psychometric Considerations

Spoken language outcome monitoring cannot occur in the ways it is recommended if the tools or methods required to do so do not exist. Recommendations for spoken language outcome monitoring have made explicit the multiple purposes spoken language outcome monitoring should accomplish. These include:

1. Provide “objective data about the individual rate of [...] development” (JCIH, 2013, p.p. e1334). To accomplish this purpose, a procedure must involve tests that are sensitive to change over time.

2. “Ensure that the child makes appropriate progress toward expected developmental milestones [...] therefore progress monitoring should be done with instruments that are norm-referenced” (JCIH, 2013, p.p. e1334). This recommendation requires the use of tests that are not only norm-referenced but
are able to predict whether a child is on a path towards age-appropriate spoken language outcomes.

3. “Guide [families’ and service providers’] decision making” (JCIH, 2013, p.p. e1334). Therefore, the results from a spoken language outcome monitoring procedure must provide clinically relevant information for individual children.

4. “[Analyze the] quality of the [EHDI Program…] for determining whether the quality, frequency, and intensity of service is sufficient” and “Develop a collaborative sharing network capable of collecting developmental data for […] the EHDI database” (JCIH, 2013, p.p. e1334). These two purposes require using tests within an EHDI program whose scores can be appropriately compared. This also requires ensuring that tests used by different EHDI programs can be appropriately compared as well.

Accomplishing regular monitoring of young children every 6 months requires tools that cover the breadth of ages serviced by an EHDI program. Deciding whether or not a child is performing within expectations requires tools that broadly capture spoken language abilities and report normative data. Additionally, we need to be assured that the tool is sensitive to the linguistic differences we might expect between children who are DHH and children who do not have hearing loss. If it is not, we cannot be assured that scores falling within normal limits reflect typical language skills or are an artefact of an insensitive tool. Deciding whether a child is making progress requires a tool that is sensitive to change over time and making progress towards expected developmental milestones requires a tool that also predicts future language development. In order to evaluate program effectiveness, the spoken language results must be comparable amongst
children to identify whether some children receiving some types of intervention are making better (or worse) progress than others. Scores derived from norm-referenced tests are specific to the normative sample and are not easily compared across different tests. If children are evaluated using different tests, it becomes difficult to identify whether differences in scores are due to differences in children’s performance or due to differences in test samples. For example, performance at the 16th percentile on Test A is not necessarily better/worse than performance at the 8th percentile on Test B due to considerations around norming sample, sensitivity/specificity, and measurement error. This issue is exacerbated in the recommendations to develop national EHDI databases to evaluate outcomes – not only do test scores need to be comparable across children within a program, but also between programs.

As a construct, spoken language is complex and represents the intersection of many domains including phonology, vocabulary, (morpho)syntax, and pragmatics. These domains represent latent traits that cannot be directly observed (Baylor et al., 2011). Spoken language tests, particularly norm-referenced tests, are predominately designed using tenets of Classical Test Theory (Baylor et al., 2011; Daub et al., 2019), which presupposes that the person’s observed score on an assessment is a quantification of the latent construct and measurement error. By measuring the latent construct through a large number of items, Classical Test Theory aims to reduce measurement error and increase the precision with which the latent construct is measured. In order to do so, tests designed under Classical Test Theory require many items intended to measure the same construct so that measurement error can be estimated and controlled. In this way, Classical Test Theory assumes that all items on a test are equally good representations of the latent
construct. Under the assumptions of Classical Test Theory, the inferences that can be made on a test result are limited to using total scores. When tests are composed only of items that are intended to robustly measure the same construct, they immediately become inappropriate for many of the purposes recommended by the JCIH (Daub et al., 2019). Measuring change, for instance, is classically thought to be antithetical to norm-referenced testing (Daub et al., 2019). Norm-referenced testing requires obtaining stable estimates of spoken language ability such that total scores can be reliably compared to normative distributions. In contrast, measuring change requires identifying whether subtle changes in ability have occurred. This is similarly the case with using norm-referenced tests to inform intervention planning. Because norm-referenced tests are classically designed to provide stable estimates of ability, they *broadly* sample the skills within a spoken language construct. Total scores are insufficient to indicate which specific abilities within the broader spoken language should be targeted in intervention.

Certainly, the JCIH recommendations are at odds with how individual spoken language tests are designed and it is not possible for a singular, norm-referenced test to accomplish all the goals spoken language outcome monitoring are intended to fulfill. This is not to suggest that the JCIH recommends using a singular test to accomplish all of the purposes of spoken language outcome monitoring. The JCIH recommendations are explicit in their recognition that EHDI programs should develop “a standard assessment battery” (JCIH, 2013, p. e1334) to accomplish spoken language outcome monitoring. However, just as individual spoken language tests require evidence that they are equipped to fulfill their intended purposes, so too do assessment batteries. Some research has used batteries to derive composite scores to measure spoken language in children who are
DHH over time (e.g., Tomblin et al., 2015), but there is no peer-reviewed guidance on which domains of spoken language, measured using which tests, should be included in a battery. Nor is there guidance regarding which recommendations are best supported by which tools. There is similarly no evidence that any assessment battery currently in use by an EHDI program results in data that are statistically equipped to address the diverse purposes spoken language outcome monitoring procedures are intended to fulfill.

1.5 Spoken Language Outcome Monitoring: Implementation Considerations

In addition to the challenge of developing an assessment battery in the absence of evidence of which tests are capable of fulfilling different components of JCIH recommendations, EHDI programs are required to ensure that the assessment procedures are (a) accurately used in clinical practice, (b) used in a way that appropriately informs clinical and program decision-making, and (c) are consistently used for all children accessing EHDI services. Incorporating numerous tests and forms of assessment may be appropriate to overcome some of the psychometric limitations described above, but EHDI programs also need to be mindful of the implementation implications of doing so.

Traditional research dissemination techniques (e.g., publication in peer-reviewed journals) rarely result in changes to clinical practice (e.g., Bauer et al., 2015; Canadian Institutes of Health Research, 2016; Eccles & Mittman, 2006). As increasing emphasis has been placed on the importance of using evidence-informed practices within healthcare, the field of Implementation Science has emerged to study the methods and strategies that effectively facilitate the regular use of evidence-based practice by practitioners and policymakers (Eccles & Mittman, 2006). As a field, Implementation
Science recognizes the complexity of factors that influence whether an individual knowledge user is likely to use new evidence in their clinical practice. Numerous theories, models, and frameworks have emerged within Implementation Science to understand implementation and support the development and evaluation of *implementation interventions* – theoretically informed strategies to support the use of research evidence (Bauer et al., 2015). Central to all theories of Implementation Science is recognition that the adoption of a new evidence-based practice is influenced by more than the quality of the evidence-based innovation itself (e.g., Atkins et al., 2017; Logan & Graham, 1998; Rogers, 2003) and successful implementation is a precursor to the improved outcomes a new evidence-based practice is expected to produce (e.g., Logan & Graham, 1998).

As it relates to spoken language outcome monitoring in EHDI programs, in order for any assessment battery to support decision-making for individual children and programs, that procedure has to be appropriately implemented across the entire EHDI program. As is the case with evidence to guide developing psychometrically appropriate assessment batteries, there is similarly no evidence documenting whether these batteries are feasible to use in clinical practice. Nor is there evidence surrounding which implementation interventions result in *accurate, sustained* uptake of assessment batteries within EHDI programs and whether data that result from outcome monitoring can be used in ways that support service providers, families, and program managers with decision-making.
Spoken language outcome monitoring, therefore, is easier said than done. EHDI programs are faced with the immense task of accomplishing spoken language outcome monitoring in the context of little evidence pointing to assessment approaches that are able to provide data that can fulfill multiple assessment purposes, little evidence that these approaches can be accurately implemented in EHDI programs, and little evidence that these approaches result in improved decision-making or outcomes for children who are DHH.

1.6 Spoken Language Outcome Monitoring in The Ontario Infant Hearing Program

Although federally funded, Canadian healthcare is provincially mandated (Health Canada, 2012) to allow provinces to respond to provincial priorities and needs. Of interest to the present dissertation is the Ontario Infant Hearing Program (IHP). The IHP was implemented in 2001 and is one of only a few Canadian EHDI programs to regularly provide adequate services for children who are DHH and regularly adhere to JCIH service benchmarks, such as timely screening, identification, and amplification (Canadian Infant Hearing Task Force, 2014; 2019). Since inception, the IHP has routinely consulted with researchers in the National Centre for Audiology at the University of Western Ontario to develop evidence-based audiological protocols, including audiological outcome monitoring, and hearing aid prescription and verification, which have been adopted internationally (e.g., Bagatto et al., 2005; 2011; Scollie et al., 2005).

Since its implementation, the IHP has closely adhered to JCIH recommendations, with slight modifications in timing recommendations to accommodate for differences in the Canadian context. When the JCIH published their recommendations to include
spoken language outcome monitoring in 2007, the IHP began to recommend administration of the *Preschool-Language Scale, 4th edition* (PLS-4; Zimmerman, Steiner & Pond, 2002) every six months for children in the IHP who were learning a spoken language. In the IHP context, speech-language pathologists (SLPs) working within another provincially funded program, the Ontario Preschool Speech and Language Program, were tasked with conducting this spoken language outcome monitoring using the PLS-4. The first cohort for which province-wide monitoring using the PLS-4 was expected was children who are DHH born in 2009.

In 2014, a series of events occurred within the ongoing partnership between the IHP and our research team in Western’s National Centre for Audiology, serendipitously creating the opportunity for the work described in this dissertation. As the inaugural birth cohort of the PLS-4 spoken language outcome monitoring procedure were discharged to school-aged services, the IHP was interested in evaluating the spoken language outcomes of this and a second birth cohort from 2011, and I became involved in analyzing the program’s outcomes (Daub, 2016; Daub et al., 2017). During these analyses, it became clear that there was substantially more missing data than we might have otherwise expected. At the same time, there were changes in the availability of the PLS-4. With the publication of the *Preschool Language Scale, 5th edition* (PLS-5; Zimmerman & Zimmerman, 2011) in 2011, the PLS-4 fell out of print and it was no longer possible for regions in the IHP to purchase test forms for this edition. This motivated the IHP to re-evaluate the outcome monitoring procedure, with the initial intention to replace the PLS-4 with the PLS-5. While plans were underway to replace the PLS-4 with the PLS-5, concerns were raised by regional coordinators and frontline SLPs about the rationale for
selecting the PLS-5. In the absence of evidence supporting the choice of the PLS-5, SLPs were concerned because they reported that the PLS-4/5 did not enhance their clinical practice, it was a time-intensive test to administer, and rumors were circulating that the PLS-4/5 was not a valid test.

The need for a new way to monitor outcomes, in conjunction with SLPs’ concerns with the PLS-4/5 and my discovery that data were substantially missing from the database, motivated the IHP to re-evaluate their spoken language outcome monitoring procedure, with careful attention to the psychometric properties of the final procedure and whether the new procedure could be implemented in a way that would minimize missing data. The following series of studies were designed to facilitate the development of a spoken language outcome monitoring protocol that fulfills JCIH recommendations in a clinically feasible way, with recognition that the recommendations must be designed in a way that supports appropriate uptake of by frontline clinicians.

1.7 The Present Studies: Purpose & Motivation

This dissertation is comprised of four manuscripts that explore the assumptions underlying spoken language outcome monitoring with the primary aim of developing a spoken language outcome monitoring protocol for the IHP. At its core, this dissertation aims to answer whether a spoken language outcome monitoring procedure can be designed and implemented in a way that results in data that can be validly used by different stakeholders for a variety of purposes.

Chapter 2 is a Viewpoint manuscript that considers best-practice definitions of validity, discusses ways in which validity has been misapplied in speech-language pathology, and provides recommendations for addressing these issues. Issues surrounding
validity are necessary to consider throughout the subsequent chapters, which were
designed to develop and evaluate a spoken language outcome monitoring procedure for
the IHP.

Chapter 3 outlines the process I engaged in with the Ontario IHP to develop a
proposed assessment battery, including the evidence considered in the design of a spoken
language outcome monitoring procedure and a candidate process. Briefly, this process
included engaging with different stakeholders (program managers, SLPs, audiologists) to
clarify the intended decisions the IHP planned to make using spoken language outcome
data and reviewing the peer-reviewed literature to identify tests that were
psychometrically equipped to support these decisions.

Through this process of developing the recommendations, managers in the
Ontario IHP agreed to pilot the recommended procedure to identify whether the
assessment process resulted in data that could be usable for their purposes, as well as
collect data about SLPs’ perceptions of that procedures, which could inform future
implementation efforts. Chapter 4 describes this pilot study, and the results of feasibility
and usability analyses used to inform the IHP’s decision as to whether they would
implement the process province-wide.

In developing recommendations for the spoken language outcome monitoring
procedure summarized in Chapter 3, I was unable to identify a suitable test of early vocal
development to recommend to the IHP, despite vocal development being an early
predictor of later spoken language outcomes. Chapter 5 reports on a survey study,
informed by Messick’s conceptual validity framework, that was designed to identify
SLPs’ assessment priorities and barriers to vocal development assessment.
Together, this dissertation is comprised of four papers designed to address the gaps in evidence required to accomplish spoken language outcome monitoring. This dissertation research was conducted with the expectation that completing this body of work would not only support the IHP specifically in developing a spoken language outcome monitoring procedure, but EHDI programs more broadly by documenting the challenges I encountered and solutions I developed in trying to fulfill JCIH recommendations.
References


Chapter 2


In order to collect objective information about clients, speech-language pathologists (SLPs) rely on a variety of different tools. Selecting, interpreting, and integrating these sources of information with other sources of evidence (i.e., client preferences and clinical expertise; Dollaghan, 2004) is a complex task that requires expertise in multiple areas, including typical performance, pathology, research methodology, and statistics/psychometrics. Test results are one important source of assessment information and can be used for various purposes including to determine eligibility for services, evaluate treatment outcomes, and determine when to discharge (McLeod & Baker, 2017; Paul & Norbury, 2012). Given the importance of test scores for clinical decision making, it is imperative that our profession is equipped with adequate tests, and that speech-language pathologists (SLPs) are adequately equipped with the resources and support to use tests validly.

Despite the importance of tests, there is mounting evidence that they are used in inappropriate ways throughout the profession, and this may be particularly true of norm-referenced tests. Kerr, Guildford and Kay-Raining Bird (2003) surveyed 144 certified SLPs in Canada regarding their use of norm-referenced tests and self-reported psychometric knowledge. Despite having an average of 12 years of experience, SLPs reported feeling only “somewhat confident” that their psychometric knowledge allowed them to evaluate tests adequately (Kerr et al., 2003). Additionally, not all SLPs were able to identify the reasons why classically defined misuses were inappropriate. Even in cases
where SLPs did correctly identify a misuse of tests, they still reported using standardized tests in ways that are classically described as inappropriate (e.g., using results to select therapy goals, see McCauley & Swisher, 1984b for a discussion; Kerr et al., 2003). In other work evaluating the frequency of test use, Betz and colleagues found that the psychometric properties of tests were not correlated with their frequency of use (Betz, Eickhoff & Sullivan, 2013). Issues surrounding test misuse also exist in the published literature, with many researchers using inappropriate testing practices in research studies (Nitido & Plante, 2020).

It is important that we understand the reasons for test misuse and identify ways to improve how SLPs use tests for assessment because test misuse has costly consequences for both SLPs and clients. For SLPs, there are the monetary costs of purchasing the test as well as the cost of time spent administering, scoring, and interpreting test results. Clients also spend time and money on assessments, but additional costs for them include the risks associated with being discharged from therapy, modifications to their intervention or educational plans, and potential (mis)diagnosis. When resources are sub-optimally allocated, costs are also passed on to insurers, funders (e.g., educational departments), and taxpayers. Daub et al. (2019) have argued that norm-referenced test misuse is a complex problem, influenced not only by SLPs’ psychometric knowledge, but also by a lack of consideration of the clinical perspective in norm-referenced test development. Addressing these issues requires work on all fronts: improving future and practicing SLPs’ knowledge, improving the design of tests, and facilitating collaborative research that incorporates both clinical and research expertise.
As a field, however, we lack an agreed upon framework to ground these conversations and begin the work of improving assessment practices. Although there has been peer-reviewed research on issues of norm-referenced assessment and test misuse, as well as systematic reviews that critically appraise the psychometric properties used by various evaluation tools (Denman et al., 2017; Flipsen & Ogiela, 2015; Plante & Vance, 1994), none make explicit their underlying framework and beliefs about validity. Additionally, there is little discussion of how these concepts might apply to other types of tests and the conditions under which SLPs should collect different evidence to inform clinical decisions. The concept of validity is important for our field to tackle because, as we will discuss, appropriate evidence cannot be dissociated from appropriate test use. That is, a SLP must understand the ways in which tests are limited in order to draw appropriate conclusions about a client’s performance.

In this viewpoint, we discuss the conceptual framework of validity originated by Samuel Messick’s (1993) influential proposal that validity is a unified concept where score meanings and decisions are the object of validation (1993). This framework is a way of thinking critically about tests and the ways in which they are, or are not, equipped to support clinical decisions (see Figure 1). This perspective has been adopted by the American Educational Research Association (AERA)’s Standards for Educational & Psychological Testing (AERA, APA & NCME, 2014) and is well suited to improve test development, assessment training (specifically, testing), and clinical practice in speech-language pathology. We have chosen the framework adopted by Standards (2014) because it is the reference standard by which psychological and educational tests are developed. Additionally, Standards provides recommendations for the responsibilities
and roles of test developers and test users. The concepts underlying the framework apply to all assessment situations in which tests are used but are most immediately applicable to standardized norm-referenced testing situations, as norm-referenced tests are often accompanied by detailed statistical and psychometric evidence (Standards, 2014).

Figure 2-1: The conceptual validity framework

What follows is a description of the conceptual framework and its components as they relate to speech-language pathology. We also consider the framework’s applicability and utility and present anticipated costs and benefits of adopting this framework. Ultimately, we argue that the conceptual framework simplifies the process of appraising validity evidence by helping SLPs focus in on only the evidence that is relevant for the decisions they want to make based on test results. We conclude with recommendations for how SLPs and other stakeholders (e.g., professional organizations, university departments, test developers) can use the framework to address issues of test misuse. Suggestions for supporting implementation of the framework into practice are also provided.
2.1 Underlying Concepts of the Validity Framework

Three key concepts are fundamental to SLPs’ understanding of validity: (a) validation refers to decisions that are made, not tests themselves, (b) collecting and evaluating validity evidence is an iterative process; and (c) there are types of evidence, not types of validity. These concepts are described next.

Validation refers to decisions, not tests. The process of determining whether a test is, or is not, appropriate for different decisions has historically been referred to as test validation, but this term is a misnomer. Tests are never (in)valid - decisions are (cf. Messick, 1993). Rather than referring to the empirical appraisal of a test as test validation, the profession would be better served by referring to this process as collecting validity evidence or decision validity. Tests are measurement tools that SLPs use to make decisions about an individual, and it is these decisions (not the tests) that can be (in)valid. Rather than collect evidence that a test possesses a certain amount of validity, the goal of collecting validity evidence is for test developers to appraise empirical evidence that identifies whether a specific set of decisions are appropriate under a specific set of circumstances. A key advantage of applying the conceptual validity framework described in Standards is that it places decisions at the heart of the validation process and links evidence with decisions, rather than requiring a single set of criteria that must be fulfilled in all circumstances.

Collecting validity evidence is an iterative, and ongoing, process. When validity evidence is first collected, developers work to generate evidence that will support a specific set of decisions. As a test is used in practice, however, SLPs will likely want to use that test for more than one purpose. For example, they may want to confirm the
presence of a disorder and measure change during an intervention period. The discrepancies between the decisions a test was developed to support and the ways in which SLPs will want to use that test have implications for all stakeholders. This includes not only test developers and researchers who can collect validity evidence to support new decisions, but also SLPs who must appraise validity evidence for every new decision they want to make. Sometimes, new validity evidence may be reported in the peer-reviewed literature after a test has been initially published which requires SLPs to evaluate new evidence.

There are types of evidence, not validity. As validity refers to the appropriateness of the decisions being made, there are not types of validity, only types of evidence. Historically, validity has been described in terms of types (e.g., face, content, construct) and in our professional vernacular, we tend to talk about whether a test possesses these elements of validity. This practice appears to be a legacy from earlier test development recommendations that described different types of validity evidence to consider during test development rather than thinking about a single gold standard of validity evidence that all tests must report (Cronbach & Meehl, 1955). However, the connection between validity and decision making was not made explicit in these earlier recommendations, which lacked a practical framework for appraising validity evidence and integrating sources of evidence with types of decisions. Focusing on whether a test does, or does not, have a type of validity is misleading because not all decisions require all types of evidence, and a test can meet the mark on all psychometric checklists and still be inappropriate for some decisions. Similarly, a test can report weak (or absent) evidence and still be appropriate for some decisions. Tests can be more, or less, appropriate for
certain decisions depending on how they were designed, and different decisions require
different types of evidence to support them.

Evaluating and integrating evidence that supports or refutes different clinical
decisions lies at the heart of modern validation research (Kane, 2013; Messick, 1995).
Evaluating the evidence requires test developers to carefully study whether the test
performs according to theoretical predictions. This results in long and statistically dense
descriptions of many sources of validity evidence that may not be immediately applicable
to the decisions SLPs want to make. Rather than looking for tests that possess all types of
validity, SLPs should look for tests that possess evidence that is relevant to the decisions
they are planning to make.

2.2 The Validity Framework

Using this conceptual validity framework, SLPs can identify whether the
decisions they make based on test scores are valid. There are three steps SLPs must use to
determine whether a decision is valid: (a) identify the decision they will make using test
scores, (b) identify what evidence they need to justify that decision, and (c) evaluate the
strength of the evidence (Kane, 2013). The first two steps can be completed without
reading or purchasing a test, and the final step requires that the SLP evaluate whether the
test manual presents adequate evidence to support the intended decision. Each step is
described in detail below.

2.2.1 Validity Framework Step 1: Identify the Clinical Decisions

The first step in applying the validity framework is for the SLP to articulate the
intended decision they plan to make using a test score. Quite simply, this step requires
SLPs to decide what they will use a test score for before selecting or administering a test. Intended decisions might include answering clinical questions such as “Are the client’s abilities lower than their peers?”, “Has the client made significant progress in therapy?” and “How severe is the client’s impairment?”. Specifying a clear and clinically relevant intended decision is essential because it will guide SLPs in selecting appropriate assessment tools. This is important as “it is typically not the case that the psychometric characteristics of a test will be optimal for multiple diagnostic purposes” (Peña et al., 2006, p. 253). Therefore, SLPs need to have access to a variety of tools to serve their different intended decisions. These tools might be norm-referenced tests for which sufficient validity evidence has been collected, or they may be other forms of criterion-referenced assessment (Betz, Eickhoff & Sullivan, 2013; Fulcher-Rood, Castilla-Earls & Higginbotham, 2018; McCauley & Swisher, 1984b), which can be more sensitive than norm-referenced tests for some decisions (e.g., measuring individual change in therapy).

Standards (2014) also highlights the fact that consequences influence a decision’s validity. Example consequences include the time and costs associated with testing, the use of test scores to grant or deny services, and the use of test scores to diagnose a disorder - which may cause social stigma or emotional distress. Standards calls for SLPs to consider the consequences of their assessment decisions on a case-by-case basis. More specifically, SLPs must consider whether the test they select reports validity evidence that is compelling enough to support the clinical decisions they intend to make. SLPs and test developers must therefore be aware of the consequences that their assessment findings might carry and take steps to minimize negative or unintended consequences. SLPs have an ethical obligation to make sure that the evidence they use to make
decisions is appropriate and balances proper test selection, use, and interpretation (Palmer, 2009). In order to fulfill these obligations, SLPs must be equipped with the skills to evaluate psychometric evidence in the context of their intended decisions, and in the context of the consequences these decisions will have. Just as what is (and is not) weak, adequate, or strong validity evidence is contextualized within a SLPs’ intended decisions, the strength of validity evidence is also contextualized within the consequences of the decision (Downing, 2003; Messick, 1995a).

2.2.2 Validity Framework Step 2: Identifying What Evidence is Needed to Support the Decision

Once SLPs have clearly stated what they intend to use a test for, they next need to work backwards and articulate what evidence they would need to be sure that the test is appropriate for answering their clinical question(s). Examiner’s manuals report many different types of evidence, some of which are statistical, and some of which are argumentative. For instance, factor analyses, correlations, and classification accuracy can be considered evidence, but so too can descriptions of how the test was developed. As a starting point, if a SLP knows the decision they want to make, they can review the examiner’s manual to find and evaluate key pieces of evidence without becoming bogged down in details that aren’t relevant for their intended decision(s). For instance, if a SLP wants to use a test for diagnostic purposes, it doesn’t matter whether the test is sensitive to change over time. Of course, some types of evidence apply to all decisions and must be considered in all cases. For instance, SLPs should always be convinced that a test accurately measures the skill it claims to measure, and that the test format (e.g., direct
response, interview, parent report) can authentically capture the underlying skill being measured.

Figures 2-2 to 2-4 outline which decisions require which evidence, and how the presence/absence of this evidence should influence whether SLPs use a test. Tables 2-1 to 2-3 expand on these figures, providing descriptions of each type of evidence, with a summary of common statistics/arguments reported in examiner’s manuals. These figures and tables are not exhaustive – instead, they represent frequently reported evidence in current examiner’s manuals. Types of evidence will certainly change over time. However, these figures and tables can provide SLPs with a starting point to guide their reading of examiner’s manuals using the validity framework. SLPs can use these resources to quickly identify what pieces of information they should be looking for, using the terminology reported in examiner’s manuals, and make decisions about whether a test is appropriate to use.

2.2.3 Validity Framework Step 3: Evaluating the Evidence

Just as certain types of evidence are relevant only for certain types of decisions, so too are certain types of statistics. Using the conceptual framework, SLPs do not need to look for, and evaluate, all statistical evidence presented in examiner’s manual. If SLPs have carefully identified their decisions, and the evidence needed to make their decisions, they can focus their appraisal on only the relevant statistics. A key problem with the traditional approach, which applies a standard set of psychometric criteria to all tests and all decisions, is that it encourages SLPs to dismiss the value of a test when certain evidence is missing or weak. The traditional approach also fails to support SLPs in
Figure 2-2: Diagnosing disorders decision tree
Figure 2-3: Measuring progress decision tree
Figure 2-4: Determining severity decision tree

1. Can I use the test to determine severity?
   - No
   - Is the test sensitive to differences in impairment levels?
     - No
     - Is the comparison sample appropriate?
       - Yes: The test is likely appropriate for determining severity
       - No: Does the test provide consistent estimates of ability?
         - Yes: The test is likely appropriate for determining severity
         - No: Does the test broadly measure the skills affected by a disorder?
           - Yes: The test is likely appropriate for determining severity
           - No: The test is inappropriate
Table 2-1: Considerations for determining the presence/absence of disorder

<table>
<thead>
<tr>
<th>Type of Evidence</th>
<th>Source(s) of Validity Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>The test accurately classifies persons with, and without, a disorder.</td>
<td><strong>Diagnostic Accuracy</strong>: Can be statistically represented a number of ways. At its simplest level, studies looking at diagnostic accuracy administer the new test to a group of people who are known to have a disorder and a group of people who are known to not have a disorder. Test developers then evaluate how often the test sorts people correctly based on their scores to derive a series of statistics representing the percentage of time the test accurately classified people. “Sorting” is accomplished by picking a cut-off score and defining people who score lower than that cut-off as having a disorder, and those who score higher as not having that disorder. Different tests are more, or less, accurate at different cut-scores and test developers should look at diagnostic accuracy at different scores to find the cut-off that is the most accurate for the test (Greenslade, Plante &amp; Vance, 2009).</td>
</tr>
<tr>
<td></td>
<td><strong>Sensitivity &amp; Specificity</strong>: Sensitivity refers to the percentage of people with a condition who will be identified as having that condition (or, achieve a certain score) using that test. Specificity refers to the percentage of people without a condition who the test will correctly identify as not having that condition. Sensitivity and specificity are based on the absolute accuracy of the test under testing conditions (where 50% of people had a disorder, and 50% of people did not have a disorder).</td>
</tr>
<tr>
<td></td>
<td><strong>Negative &amp; Positive Likelihood Ratios</strong>: Similar to sensitivity and specificity, negative and positive likelihood ratios estimate the likelihood that someone does, or does not, have a disorder. Critically, however, negative and positive likelihood ratios account for difference in testing rates of a disorder (Dollaghan, 2007). That is, in clinical contexts where a disorder is rare, a SLP is likely to see very few people who truly have a disorder and they are more likely to see people who score above versus below the cut-off. Both sensitivity/specificity and negative/positive likelihood ratios are important for tests to report, and the information from each statistic should be considered (Lange &amp; Lippa, 2017).</td>
</tr>
<tr>
<td></td>
<td><strong>Evidence based on content</strong>: The test questions have been reviewed by content experts and all domains relevant to disorder are covered.</td>
</tr>
<tr>
<td></td>
<td><strong>Evidence based on response process</strong>: How a person answers questions on a test (e.g., pointing, naming words) is appropriate for the skill being measured.</td>
</tr>
<tr>
<td></td>
<td><strong>Evidence based on relations with other variables</strong>: Peoples’ scores on the test are associated with scores on other tests/measures of a similar ability (ideally, a gold standard) and are not associated with scores on tests/measures of other skills that are not relate to the underlying skill. These studies typically have a group of participants complete two (or more) tests and report correlations between the test scores.</td>
</tr>
<tr>
<td></td>
<td><strong>Evidence based on structure</strong>: This evidence is commonly reported as factor analyses or structural equation modelling. A test should contain as many “factors” or “latent variables” as the skills it claims to measure. Each factor should consist of all the items related to the skill. If the test is measuring a single skill (e.g., receptive vocabulary), it should contain one factor and all items should correlate with one another. If the test measures multiple skills (e.g., an omnibus language test) then there should be multiple factors with items relating to their appropriate factor.</td>
</tr>
</tbody>
</table>
The test provides consistent estimates of a person’s ability.

**Administration properties:** The examiner’s manual provides sufficient detail that the SLP is confident they are administering the test in the same way as test developers.

**Reliability evidence:** Reliability can take many forms and is often reported in the form of correlations. Deciding that a correlation is high enough to be considered reliable is subjective, but tests aim to have a minimum reliability (across types) of 0.80 or higher (Terwee, 2011)

**Test-retest reliability:** Typically, a correlation between one person’s scores on the same test taken on different days. Test-retest reliability measures how stable a test score is over time.

**Alternate/parallel forms:** Some tests will develop alternate forms that are intended to measure the same skill. These alternate forms help with re-assessment because a person will be asked different questions and cannot memorize correct responses. In these situations, there should be evidence that scores on both forms are highly correlated with one another.

**Inter-rater reliability:** Typically, a correlation between a person’s test score when they are evaluated by two different examiners.

**Internal consistency:** A number of metrics indicating that a person’s responses to different questions on the same test are consistent with one another and the test is reliably measuring the appropriate skill. May be reported as a Cronbach’s Alpha (Cronbach, 1951), or Revell’s Omega (McNeish, 2017; Zinbarg et al., 2005), value. Alpha or Omega values greater than 0.80 are generally accepted as appropriate in speech-language pathology.

**Standard Error of Measurement:** The variability of a measured score around the participant’s true score. All tests are associated with some degree of error, and standard error of measurement provides a range of possible scores around the score a client receives that could represent the client’s ability. For example, if a client receives a standard score of 50, with a standard error of measurement of 5, their true score is somewhere between 45 and 55. This is important to consider when cut scores are used in diagnostics. If the standard error of measurement includes scores that are both above and below the cut score, it is unclear whether the client meets criteria for a diagnosis.

Measurement error should be specified and the smaller the error is, the more confidence one can place on the actual test score.
### Table 2-2: Considerations for determining change in skills

<table>
<thead>
<tr>
<th>Type of Evidence</th>
<th>Source(s) of Validity Evidence</th>
</tr>
</thead>
</table>
| The test is sensitive to change in ability. | **Criterion-referenced tests** *(McCaughey & Swisher, 1984b)*: Criterion-referenced tests are not typically standardized, and often measure very specific skills rather than a wider range of abilities. Criterion-referenced tests should have questions that are very specific to the skill being addressed in therapy (e.g., picture-naming).  
  
  **Item Response Theory derived analyses for tests that are also norm-referenced**: Some analyses and scores reported in norm-referenced tests have been specifically developed using an analytic set such as *Rasch or Latent Trait’s Models* that measure how much ability a client has in the skill being tested, instead of the client’s performance compared to their peers. These analyses can be used to create *Growth Scale Values*, which are scores that measure whether a client has made progress, fallen behind, or not developed new skills, relative to their previous score. These scores are more sensitive to change in skills following intervention than standard scores and percentile ranks *(Daub et al., 2017)*. |
| The test provides consistent estimates of a person’s ability. | The validity evidence summarized in Table 2-1 broadly applies to using a test for determining change in skills. This evidence includes evidence for reliability of all types (i.e., test-retest, internal consistency, inter-rater, alternate/parallel forms) and is evaluated using the same criteria (correlations greater than 0.80).  
  
  In measuring change in skills, high test-retest reliability and small standard errors of measurement, are very important. This is because if a test has low test-retest reliability, this means that large differences in scores over time can be due to error and not due to growth. Similarly, large standard errors of measurement (large uncertainty around a person’s true test score) means that a person needs to score very differently in order to decide that true change has occurred. |
| The test measures the ability being targeted in therapy. | The sources of evidence described in Table 2-1 broadly apply. This includes evidence based on content, response process, and relations with other variables. Specific to measuring change, this evidence should convincingly demonstrate that the test contains items relevant to the ability being targeted in therapy. |
| The test sample represents the person being tested. | Examiner’s manuals should provide a detailed report of the demographic characteristics of the normative sample. Comments on the similarity of these demographic characteristics to a broader population (e.g., using census data) will be useful to determine if the normative sample provides a reasonable representation of the examinee’s peer group. |
### Considerations for determining severity/level of function

<table>
<thead>
<tr>
<th>Type of Evidence</th>
<th>Source(s) of Validity Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>The test is sensitive to differences in impairment levels</td>
<td><em>Classification accuracy:</em> This could include statistical evidence similar to diagnostic accuracy statistics, where the test attempts to sort clients into different severity categories (see Table 2-1). Instead of sensitivity and specificity to the presence or absence of a disorder, a test should report sensitivity/specificity and negative/positive likelihood ratios for the test’s ability to classify people into different severity levels (i.e., mild, mild-moderate, severe, profound).</td>
</tr>
<tr>
<td>The comparison sample is appropriate.</td>
<td><em>Test scores are linearly predicted by severity ratings:</em> This can be evidenced by correlations between levels of severity (established using a gold standard) and test scores. <em>The normative sample includes clients with a wide range of severity impairments:</em> In some cases, it may be appropriate to compare a person’s score to individuals without impairment because their impairment relates to a difference rather than disorder. However, it is important that a normative sample includes individuals with a wide range of impairment severity.</td>
</tr>
<tr>
<td>The test provides consistent estimates of a person’s ability.</td>
<td><em>A comparison sample is derived entirely of clients with disorders and captures the range of impairment severity:</em> When a client’s skills are speculated to be fundamentally different from their peers without an impairment, comparing their performance to healthy controls is not always appropriate. In this case, a separate comparison sample might be appropriate.</td>
</tr>
<tr>
<td>The test broadly measures the skills that are affected by a disorder area.</td>
<td><em>Which of the two sources of validity evidence is appropriate depends on the clinical population the SLP is serving.</em> The validity evidence summarized in Table 2-1 broadly apply to using a test for determining severity/level of functioning. This evidence includes evidence for reliability of all types (i.e., test-retest, internal consistency, inter-rater, alternate/parallel forms) and is evaluated using the same criteria (correlations greater than 0.80). Small standard errors of measurement are particularly important so that subtle changes in scores can be sensitive to differences in severity levels. The validity evidence summarized in Table 2-1 related to test content broadly applies. This includes evidence based on content, response process, and relations with other variables. Specific to measuring severity, the test should contain items that are important to deciding whether a person has a severe impairment and the response process should be appropriate for people with a wide range of severities. In addition, tests should incorporate items related to the Activities and Participation components of the International Classification of Functioning, Disability, and Health framework (World Health Organization, 2001) in order to determine whether a person’s health condition is impairing their daily life.</td>
</tr>
</tbody>
</table>

---
making decisions when a test has inadequate levels of evidence in some, but not all, areas. For instance, what should a SLP do if a test’s classification accuracy is low, but there is evidence that the test is internally consistent? Using a checklist, SLPs might be inclined to view the test as “invalid” and disregard it entirely, but SLPs can take a more nuanced approach using the validity framework. If the intention is to use the test for diagnosis, and the test doesn’t accurately classify people with/without a disorder, the test should be discarded. But if the intention is to measure change, then classification accuracy is irrelevant, and the test may still be appropriate. Through the shared lens of a validity framework, communication between test developers and SLPs can support the reporting of statistical analyses that are necessary for the specific decisions SLPs intend to make and the development of tests and reporting practices that are useful and meaningful for clinical practice.

2.2.4 Integrating the Validity Framework in Practice

Using this framework requires asking a series of questions, the first of which is never “is this test valid?”. Tests are never categorically valid or invalid, and there is no single piece of evidence that can be used to claim one tool is better or more valid than another.

Although this viewpoint discusses the three components of the validity framework as distinct, the process of using this framework is not linear, and is similar to decision making in evidence-based practice in which three sources of evidence, client preferences, clinical expertise, and the best available research evidence, are integrated to make decisions (Dollaghan, 2004, 2007; Sackett, Rosenberg, Gray, Haynes & Richardson, 1996). This process is not intended to disregard clinical expertise or client
preferences but rather it highlights the importance of carefully integrating each source of evidence to inform clinical decisions. The same principles apply to test selection and application of the validity framework.

Viewing validity through the lens of evidence-based practice requires SLPs to consider a client’s preferences and readiness to engage in testing, and the consequences of testing. If a client requires a diagnosis and access to intervention services, then SLPs will likely administer a test that helps them decide whether the disorder is present. SLPs should use their clinical expertise to carefully map out what they know about both testing and the disorder in order to specify the validity evidence a test should report for decisions to be valid. SLPs should then seek out the best available validity evidence, that is, the test that is best equipped statistically to answer their specific question about a specific client in a specific circumstance. It is also important to highlight that all tests will have limitations and all statistics are associated with some degree of error. As an example, it is unlikely that any test will correctly identify a disorder 100% of the time, even if it is otherwise an appropriate diagnostic tool. How then, do SLPs account for error? SLPs should return to their intended decision, consider the consequences of making an error, and identify methods for mitigating those consequences. Specific methods may include collecting additional assessment evidence and tempering the strength of clinical conclusions when it is appropriate or necessary.

To demonstrate this decision-making process, we next provide an overview of clinical decisions that are commonly made using norm-referenced tests, the types of evidence that are necessary to evaluate in order for a test to be useful for those decisions, and the sources of this evidence. Descriptions of appropriate validity evidence are
derived from *Standards* (2014). The examples here represent common purposes and do not exhaustively cover all decisions SLPs might make using a norm-referenced test, nor do they cover all potential decisions that *could* be validly made. Some types of validity evidence apply to multiple decisions (e.g., measurement consistency) whereas others are unique to specific decisions. First, we provide an overview of those types of evidence that are necessary to support all assessment purposes, and then a discussion of the key sources of evidence for specific purposes.

### 2.3 Example Decisions and Necessary Validity Evidence

*Generally applicable evidence.* All tests, regardless of their purpose, must provide satisfactory evidence that the test measures what it claims to, was evaluated using an appropriate sample, and consistently measures a person’s abilities. These types of evidence support an SLPs decision making because they can be assured that the test is a reliable measure of the relevant skill and is comparable to the person being tested. Tables 2-1 to 2-3 describe these types of evidence and their specific relevance to each example decision.

*Determining the presence/absence of a disorder.* Determining whether someone has a disorder is based on the notion that their performance in a particular domain is so much poorer than a typical population that they belong to a different group or population (see Table 2-1). The goal of testing for this purpose is one of classifying which group the person belongs in: someone with a disorder or someone without (Dollaghan, 2007). How this classification occurs, may be different depending on the specific nature of a disorder. Diagnosing developmental language disorders requires evidence that the child’s abilities are sufficiently below their peers and that they belong to a different group (children with
a language disorder). In this case, evidence is needed that there is a cut-off score on the test that can accurately classify whether the child belongs to a group of typically developing children, or a group of children with language disorders. In adults with acquired language disorders, however, the nature of their disorder isn’t developmental – the errors that they make aren’t made by healthy adults and therefore don’t need to be differentiated from a normal distribution. In these cases, SLPs can rely on different sources of evidence, such as the presence of clinical markers or otherwise disordered error patterns. However, the use of these clinical markers requires similar evidence that the markers themselves differentiate between persons with, and without, impairment.

Measuring change. There are different ways SLPs can define and describe change. One way to define change is for SLPs to describe either the acquisition or loss of individual skills, often in response to treatment (see Table 2-2). This can be done using informal measures such as criterion-referenced tests (which are more sensitive to change in specific skills than norm-referenced tests) or using appropriate scores from norm-referenced tests. If SLPs want to describe the acquisition or loss of individual skills using a norm-referenced test, then they require evidence that the test is sensitive to individual change over time. This can involve consideration of confidence intervals and standard error of measurement to decide that change is not due to measurement error. SLPs can also use scores derived from latent trait’s models (e.g., growth scale values; Daub et al., 2017). Growth scale values are a type of score designed to measure growth in a client’s level of ability, instead of their standing relative to their peers. Although growth scale values are limited in that they are derived from tests which often measure more skills than might have been worked on in therapy, they can be useful for measuring change in
overall ability. Change in ability can be detected by comparing growth scale values from two time points and deciding whether the difference between the two scores is greater than the confidence intervals reported in the examiner’s manual. If the second assessment score is higher than the first, and the difference is larger than the reported confidence interval, then the client has made statistically significant gains. If the second assessment score is lower than the first, and the difference is larger than the confidence interval, then the client lost skills.

Alternately, SLPs can describe change in a person’s ability relative to their peers (see Table 2-1 for validity evidence). People can catch-up, fall behind, or stay the same in their relative standing. Using a combination of cut-scores and confidence intervals, SLPs can determine whether a client has caught up, fallen behind, or stayed the same relative to their peers.

*Measuring severity or level of function.* Validity evidence surrounding classification accuracy provides SLPs with assurance that test scores can be appropriately used to determine the presence or absence of a disorder. This evidence is insufficient, however, to decide how severely the person is affected. Classification accuracy refers to whether cut-off scores on the test accurately sort people into groups: people with an impairment and people without an impairment. But the groups used in test development might not have included all levels of severity – or equally sampled from the full range of severity. That is, there might be more people in the group who have moderate levels of impairment than have severe impairments. For disorders where a normative distribution is inappropriate to describe the population (because healthy individuals do not have the same characteristics as people with the disorder) a test will require an adequate
comparison sample. This sample should contain individuals with an impairment from the full range of possible severities. Or, if estimates from some levels of severity cannot be obtained (perhaps because of difficulties for those individuals in completing the test), then this should be specified. In this case, SLPs should be cautious about using the test to measure severity in the ranges that are not included in the test sample.

For disorders where a normative distribution is appropriate to describe the impairment (because the disorder represents skills that are below age-expectations), a normative distribution is still not always sufficient for determining severity. This may appear counter-intuitive. If normative scores form a normal distribution, why are lower scores not necessarily associated with more severe impairments? The answer contains two parts. First, any score a person receives on a test is associated with some degree of measurement error. That is, due to any number of reasons, if they were to be tested on a different day or at a different time, their score would be slightly different. Because all scores are associated with error, in the absence of appropriate validity evidence, it is difficult to determine whether a person with a standard score of 50 is more impaired than a person with a standard score of 55. Second, the population used in the normative sample influences a test’s sensitivity to impairment severity (Peña, Spaulding & Plante, 2006). Normative samples rarely include people with the full range of impairment severity. Without data for how people with the full range of disorder severity perform on a test, it is difficult for a SLP to decide that a person’s score indicates more, or less, severity relative to others with a communication disorder. The types of validity evidence that are important for SLPs to consider when using a test to determine the severity of a person’s impairment are presented in Table 2-3.
2.4 Benefits of Adopting the *Standard’s* framework

There are both costs and benefits to adopting the proposed conceptual framework of validity that must be considered from multiple perspectives including professional development, client services, and test design. The costs and benefits of each perspective are presented next, but overall, we believe the benefits outweigh the costs.

*Professional Development.* Adopting a consistent validity framework for widespread use requires SLPs, including those who may be many years post-graduation, to engage with psychometrics in a new way and in order for adoption of the validity framework to be successful, significant knowledge translation and implementation efforts would be required. Fortunately, by re-centering validity discussions on decisions rather than tests, this framework can be adapted to match SLPs’ current understanding of assessment so that knowledge translation can focus on the psychometric concepts associated with specific decisions.

Adopting the framework can result in significant time-savings. Targeting specific intended decisions, one at a time, allows SLPs to focus on the hypotheses and types of evidence that are relevant for the decisions they intend to make. As a starting point, SLPs can read through the manuals of candidate tools to identify whether the decisions they want to make have been validated by the tool. If validity evidence has not been collected for this purpose, SLPs can quickly dismiss the test and move on. For tests that include relevant validity evidence, SLPs can then review the examiner’s manual in more detail to confirm the tool is appropriate.

The validity framework also lays the groundwork for those teaching future SLPs to support learning about test psychometrics in relevant and pragmatic contexts. Rather
than teaching all elements of psychometric properties together, professional coursework can emphasize the assessment purposes relevant to a clinical scenario, highlight the decisions that a student might wish to make, and teach students to understand the statistics they need to support that decision. For example, Peña, Spaulding and Plante (2006) discuss the use of tests for relative (e.g., assigning severity ratings) or absolute (e.g., diagnosing the presence or absence of a condition) purposes in language diagnostics. To increase a test’s accuracy for absolute purposes, we need tests that exclude individuals with mild impairments, but to increase a test’s accuracy for relative purposes, we need tests that include individuals with a wide range of impairment severities. Including or excluding clients with mild impairments is not inherently correct or incorrect in validity studies, and by emphasizing the validity of decisions rather than tests, Standards’ (2014) framework provides an opportunity to discuss the relevance of test samples to psychometric appropriateness.

**Improving Client Care.** The costs of adopting the Standards’ framework for clients is less clear, however, the benefits are obvious. In fact, we argue that the people who stand to benefit the most from the adoption of this framework are clients.

Strong clinical competence in validity will not only improve the quality of individual clients’ service and but also allow SLPs to better advocate on clients’ behalf. Consider, the well-documented problem of state education departments mandating arbitrary norm-referenced cut scores to determine service eligibility (Betz, Eickhoff & Sullivan, 2013; Spaulding, Szulga & Figueroa, 2012). The cut scores established by state education departments are not only inconsistent with one another (Spaulding, Szulga & Figueroa, 2012), but the emphasis on applying a single cut-off criterion is also
inconsistent with a) existing best-practice in speech-language pathology (Peña, Spaulding & Plante, 2006; Spaulding, Szulga & Figueroa, 2012), b) existing best-practice in test development (Standards, 2014) and c) federal legislation in the United States (i.e., the Individuals with Disabilities Education Act, 2004). These state-level policies ask SLPs to violate federal legislation and deny services to children who clinically meet the thresholds for service provision (Hogan, 2019). To change policy, SLPs must be equipped with a strong understanding of validity evidence and why arbitrary cut-offs are inappropriate. SLPs must also understand how policies could be improved in an equitable, evidence-based way. The benefits to professional development described earlier carry direct benefits to clients by enabling SLPs to effectively advocate against inequitable policies.

**Test Design.** Under the validity framework, evaluating validity evidence becomes more nuanced than evaluating test properties according to checklists of psychometric criteria. Using this framework, it is insufficient to argue that because a given statistic is significant, the test is appropriate, should be adopted, or is otherwise “valid”. Standards places responsibility on test developers to consider the types of decisions SLPs might make based on test results so that statistical evidence is presented for the decisions the test is versus is not equipped to support.

The validity framework can also broaden our field’s perspective on how tests may be used. We know that SLPs often use tests (particularly norm-referenced tests) to make multiple decisions (Kerr, et al., 2003). As statistical methods evolve to support validity evidence for an increasing number of decisions, test developers could collect validity evidence to explicitly demonstrate which decisions are (in)valid. This is not to say that
test developers ought to be responsible for evaluating every potential decision a SLP might make prior to publication. Collecting these data can be costly and might mean delays in test publication. Instead, under this framework, collecting validity evidence is an ongoing process because SLPs could make innumerate decisions requiring innumerate validation studies. Test-developers and SLPs need to collaborate to decide how much validity evidence is enough, how much error or uncertainty is tolerable, and which decisions should be prioritized in the validation process.

The validity framework offers exciting opportunities for researchers engaged in test development to collaborate with SLPs to understand and document the clinical decisions they want to make. Various qualitative and mixed methods techniques exist to support identifying and prioritizing the goals of SLPs (e.g., concept mapping; Kane & Trochim, 2003). Engaging SLPs in the initial stages of collecting validity evidence will ensure tests are clinically relevant and may result in tests SLPs are more interested in using and purchasing. At present, not all tests meet the standards of evidence to which we should hold them (Denman et al., 2017; Flipsen & Ogiela, 2015; Plante & Vance, 2015), but our field has seen major progress in test design when we unite our voices in calling for change (Daub et al., 2019). The validity framework and mixed-methods approaches to test design both lend themselves to integrating the framework of evidence-based practice into test design. Using the Standards’ framework, test developers can engage with SLPs and clients to identify which assessment decisions are high priority so tests can be designed to include the statistics that support these decisions. At a fundamental level, test publishing companies are motivated to design useful tests that are likely to be purchased,
and publishers stand to benefit from engaging SLPs and clients in test development using the validity framework.

Additionally, the framework’s emphasis on decisions, rather than tests, creates the possibility for test developers to re-center examiner manual discussions in ways that are more accessible to SLPs. For instance, some tests (e.g., the *Preschool Language Scale 5th ed.*; Zimmerman, Steiner & Pond, 2011) have earned poor reputations amongst SLPs as being “invalid” because sensitivity and specificity are low at cut-off scores SLPs use for diagnosis (e.g., Elleseff, 2018; Smith, 2014; ), despite its appropriate classification accuracy at other cut scores and similar accuracy to other preferred tests. Test developers can overcome these issues by making explicit the connection between the decisions the test is designed to address, and the relevant evidence. The decision trees (see Figs. 1 – 3) can be modified to produce summary documents outlining what evidence is, or is not, reported in examiner manuals. Tailoring examiner manuals towards clinical decisions is mutually beneficial as it would make it easier for SLPs to find key information and therefore more likely they would purchase tests.

2.5 Next Steps Towards Adopting the Validity Framework

The conceptual validity framework has the potential to improve test design and support clinical capacity. Clients are entitled to receive care in which high-quality tools are used in the most appropriate way. It is the clients above all others, who stand to benefit from improvements in test design and from SLPs who have strong knowledge of the framework. Thus far, we have provided an overview of the conceptual validity framework and outlined the costs and benefits of the framework to test design, professional development, and client services. In sum, we believe the benefits of using
the conceptual framework outweigh the costs, but we recognize that knowing about the framework is insufficient to change clinical practice. Research in implementation science has routinely demonstrated that moving evidence into clinical practice is an intentional, active process of identifying and overcoming barriers. Implementation requires structural, organizational, and individual efforts from multiple stakeholder groups (Greenhalgh et al., 2004; Olswang & Prelock, 2015). Professional organizations, universities, researchers in test-development, and SLPs all have different roles and responsibilities in supporting the implementation of the validity framework.

*Professional Organizations.* Professional organizations (i.e., ASHA, Speech-Audiology Canada and state/provincial regulatory bodies) are in an important position to facilitate uptake of the validity framework as they can influence both university curriculum and professional development. Their support is imperative in achieving uptake. As a first step, professional organizations can support uptake of the validity framework by developing clinical practice guidelines and recommendations. With explicit recognition of the framework’s value, diverse strategies can then be employed to support systemic change in testing practices. For instance, task forces can be developed to study how testing and assessment is currently taught in university programs, create recommendations for achieving unity, modify curriculum expectations to maintain accreditation, or inform new content for entry to practice exams to ensure clinical competency. We recommend that a task force includes SLP, test developer, and researcher representatives, so all perspectives are considered.

Professional organizations can also support SLPs in developing their skills in testing and validity knowledge through the development and distribution of educational
resources. Such resources could include newsletters, clinical practice guidelines, communications to their membership (e.g., webinars and infographics), and workshops presented at national conferences. Review of these resources could then be considered towards existing professional development requirements, preventing additional burden for SLPs. However, the availability of high-quality research evidence and educational materials is insufficient to change clinical practice (Graham et al., 2006). Active implementation efforts will be required to ensure new knowledge is successfully implemented into practice (Bauer et al., 2015; Graham et al., 2006), and research will be needed to demonstrate whether practice change occurs in response to these efforts. Measuring implementation success could be one activity of a designated task force.

**Universities.** University departments and faculty responsible for training future SLPs have a critical role to play in supporting the implementation of the validity framework in both research and practice. Graduate training programs are uniquely positioned to shape future SLPs’ understanding of validity and to encourage research trainees to evaluate how they report validity evidence. Ultimately, uniformly adopting the Standards validity framework across the profession means that some, if not all, training programs will need to modify curriculum to align with best practice. We recognize that this is a considerable undertaking, but we believe the benefits outweigh the costs. As such, we recommend faculty and departments review their curriculum to consider the way validity is currently taught.

This may prompt questions of what a revised curriculum would look like. In some circumstances, it may mean creating or modifying existing evidence-based or professional practice coursework to include coverage of the validity framework. In
others, it may mean incorporating validity and psychometric training into all courses where assessment is considered to support students in making explicit links between clinical decisions and validity evidence. Experiential learning opportunities that are currently a part of many training programs can be modified to require students to demonstrate an understanding of why they selected a test and how the statistical evidence underlying the test should influence their decision-making. Curriculum modifications may vary from institution to institution, but what should become standard expectation is that trainees graduate with the clinical competency to (a) link the decisions they make to validity evidence, (b) identify the inferential limits of different tests, (c) identify alternative sources of clinical information to overcome these limitations, and (d) articulate the consequences of, and solutions to, test misuse from a variety of stakeholder perspectives.

**Researchers & Test developers.** As previously discussed, adopting the *Standards* framework carries exciting possibilities for mixed-methods research in test design. We have discussed the importance of incorporating SLPs’ perspectives in test development and evaluation, but researchers and test developers will also play an important role in supporting practice change, especially considering their unique expertise in the psychometric appropriateness of various statistics to support specific decisions. Contributions may include research programs in implementation science dedicated to understanding the barriers and facilitators associated with implementing the framework into clinical practice, and the development of implementation interventions to overcome barriers and support stakeholders in changing practice. The scientific community also has a role to play in developing, evaluating, and modifying training resources such as
tutorials, workshops, and webinars to ensure that information is presented accurately, and the limitations of statistical analyses are clearly described.

We also recommend the research community reconsider the language used to describe the ways in which validity evidence is described and reported. Although researchers might be aware of, and use, the conceptual validity framework the link between how validity evidence should inform clinical decision making needs to be made explicit, using terminology that is easily recognized by SLPs. As previously discussed, test validation is a misnomer and perpetuates misunderstandings in validity evidence. The terms collecting validity evidence or decision validity are more appropriate. Furthermore, researchers should consider the language used to report validity studies. Peer-reviewed papers would benefit from explicit connection to the validity framework, relating the study hypotheses to clinical decisions, and linking study limitations and next steps to future clinical decisions. If tests are designed using mixed-method approaches, understanding the terminology that SLPs use to describe their decisions can be facilitated. Themes, or direct quotes that emerge from this work can be used to describe study purposes, and next steps. For instance, what is the clinical importance of a study that evaluates the correlation between a new test and an existing test? Ultimately, in order to determine whether future decisions (such as the presence or absence of a disorder) are valid, it is important to know that the test measures the underlying construct it claims to measure. If the goal is to use the test for diagnosis, next steps would include developing normative expectations and evaluating the test’s classification accuracy. We expect that SLPs’ perspectives will help guide the way results are framed in relation to clinical decisions and support readability by clinical audiences.
SLPs. Primarily, we argue that SLPs’ immediate next steps are to reconsider how they currently appraise validity evidence and use tests. We anticipate that the tables and decision trees incorporated throughout this paper can be used as a guide for SLPs wanting to implement the framework in their practice. In addition to changing practice, SLPs also have an important role in supporting the implementation of the framework across the profession. SLPs’ perspectives are of paramount importance because any implementation efforts need to be feasible and meaningful before they can influence clinical practice. As a first step, we recommend SLPs advocate for the inclusion of their perspectives in developing both tests and professional development materials. SLPs may participate by discussing areas of psychometrics where they require additional training support, or by identifying barriers and facilitators to their own professional development. Only with their involvement can educational materials and supports be tailored to SLPs’ needs.

SLPs can further support implementation efforts by sharing the importance of best testing practices with their clients. As clients are directly invested in the outcomes of testing, it is important for them to be supported in making informed decisions about the tests that are used to assess their performance. Tests should be developed to answer questions important to both SLPs and their clients, but we acknowledge that clients may be in the weakest position to advocate for changes in testing practices. Where test development can be modified to include SLPs’ perspectives, the same modifications can be made to include clients’ perspectives as well. Internationally, healthcare research is increasingly recognizing the importance of collaborating with clients as research partners (Canadian Institute of Health Research, 2011). To this end, SLPs can advocate for
clients’ inclusion in test development and connect interested clients with opportunities to be involved.

2.6 Conclusions

In this viewpoint, we have argued that adoption of the conceptual framework for test validation provides a way forward for the profession to begin improving testing practices. Adoption of the framework would provide clarity and shared terminology between SLPs and researchers to ensure the development of feasible and meaningful tests, and more appropriate use and interpretation of tests. We have recommended a series of next steps for multiple stakeholders across the profession. One key recommendation is the recognition that it is the clinical decisions that are being validated, and not tests themselves that lies at the core of the Standards’ framework. Although SLPs may not be experts in psychometrics and statistics, they are experts in clinical practice, and their inclusion in the process of test development will foster the design of tests that are more relevant to practice and easier to use. Adoption of the framework by all stakeholders will also simplify efforts to educate SLPs and trainees about psychometrics.

We expect that by improving the usability of tests, and simplifying psychometric education, testing practices will improve. Future interdisciplinary work integrating the perspectives of test developers, SLPs, and clients will evaluate the extent to which these arguments are true. Future research will also evaluate the ways in which this framework can be applied to improve test design, critical appraisals, and professional education, as well as whether implementation efforts result in changes or improvements in clinical practice.
References


Hogan, T. (Host). (2019). Lively discussion about tests with Elena Plante: Why we use them, what makes one valid, is more better, how are they different now? (Episode 7). In SeeHearSpeak. [https://www.mghihp.edu/research-research-labs-speech-and-language-sail-literacy-lab/seehearspeak-podcast](https://www.mghihp.edu/research-research-labs-speech-and-language-sail-literacy-lab/seehearspeak-podcast)


Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach’s $\alpha$, Revelle’s $\beta$, and McDonald’s $\omega_H$: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7
Chapter 3

3 Developing a Spoken Language Outcome Monitoring Procedure for a Canadian Early Hearing Detection and Intervention Program: Process and Recommendations

Early Hearing Detection and Intervention (EHDI) programs provide family centered support in the pursuit of typical language development (whether signed or spoken) for children who are deaf and hard-of-hearing (DHH; Moeller et al., 2013). For families who elect to pursue language in a spoken modality, EHDI programs have been demonstrated to improve spoken language outcomes (Ching, Day et al., 2013; Moeller, 2000; Tomblin et al., 2015). Recent research has identified that interventions provided through EHDI programs such as early amplification, high levels of audibility, and support for consistent hearing aid use, are significant predictors of eventual spoken language outcomes and growth in spoken language over time (Tomblin et al., 2015).

Comprehensive EHDI programs are gaining increasing international support, and international recommendations have been developed to guide their implementation (Moeller et al., 2013; Joint Committee on Infant Hearing, 2013, 2019). The Joint Committee on Infant Hearing (JCIH) has worked for many years to establish guidelines to ensure consistent and equitable service for children who are DHH enrolled in different EHDI programs across the United States, and their work has set a standard for EHDI programs worldwide (e.g., the Canadian Infant Hearing Task Force endorses these recommendations). One of the committee's activities has been publication of position statements summarizing the current state of the evidence in infant hearing and providing
preferred practice recommendations on early identification and intervention for children who are DHH.

Of interest to the present article are JCIH recommendations for routine outcome monitoring of children enrolled in EHDI programs, specifically the monitoring of language outcomes. Because a central aim of EHDI programs is to prevent developmental delays associated with permanent childhood hearing loss, the recommendation for routine monitoring of spoken language development (when this is the mode of communication chosen by the family) is intended to ensure that “a child’s developmental progress is comparable with his or her hearing peers” (JCIH, 2007, p. 909) and within 1 SD of their age or cognitive development on norm-referenced spoken language testing (JCIH, 2013). To meet this expectation, the JCIH recommends that policymakers, service providers, and family members use the results of routine spoken language outcome monitoring to support decision making. For instance, results from spoken language monitoring should be used to inform program evaluation and quality assurance at the program level, support comparison between EHDI programs using national databases, inform intervention planning at the level of the individual child and family, and determine whether a child is or is not meeting developmental milestones (JCIH, 2013, 2019).

However, there is no clear guidance on how EHDI programs ought to accomplish spoken language outcome monitoring, and the concept of spoken language outcome monitoring is poorly defined. Spoken language encompasses a wide range of inter-related skills, some of which a child may or may not struggle with at different ages. Nor do recommendations connect assessment purposes with tests or propose solutions to
overcome the psychometric challenges associated with defining acceptable outcomes. Identifying the intended purpose(s) of conducting routine measurement of spoken language outcomes is an essential consideration in selecting the assessment approaches and which tests to use (Daub et al., in press), because different tests may be better suited to different purposes. Furthermore, not all tests are validated to support multiple decisions (Daub et al., 2019) and some assessment purposes are at psychometric odds with one another. For instance, the appropriate composition of a normative sample changes if the test is being used for absolute purposes (i.e., determining whether a child is below age expectations) or relative purposes (determining the severity of a spoken language disorder; Peña et al., 2006). As outlined by JCIH (2007, p. 909), “the primary purpose of regular developmental monitoring is to provide valuable information to parents about the rate of their child’s development as well as programmatic feedback concerning curriculum decisions.” These two decisions (i.e., information about rate of development and programmatic feedback) imply two conflicting purposes: measurement that is sensitive to an individual child’s growth over time and measurement that is comparable between all children in a program. In speech-language pathology, it is traditionally advised to avoid measuring growth with norm-referenced tests because these tests are inherently broad, robust, and stable measures of spoken language constructs that aren’t designed to be sensitive to change in language ability (McCauley & Swisher, 1984). However, relatively new statistics (e.g., item response theory derived scores such as growth scale values) that can be used to measure change over time are increasingly being reported in norm-referenced tests, although these are not yet commonplace (Daub et al., 2017; Daub et al., 2019). Comparing results between groups of children for the
purpose of evaluating the broader EHDI program, however, requires that all children in
the program are assessed at regular intervals with a consistent measure so that norm-
referenced results can be compared.

The present project was born out of our efforts to support a Canadian EHDI
program, the Ontario Infant Hearing Program (IHP), which serves children from birth to
age 6, in developing a spoken language outcome monitoring procedure that would allow
them to fulfill best-practice recommendations. The IHP was developed in 2002 and is a
publicly funded EHDI program. The IHP provides universal newborn hearing screening
services to all babies born in Ontario and intervention services to children with
permanent hearing loss up to the age of 6 years. Spoken language development services
for children in the IHP are provided by the publicly funded Ontario Preschool Speech and
Language Program until they transition to school services, which can start as early as 3 or
4 years for those who attend junior kindergarten, but does not occur until 6 years of age
for others. The IHP provides language development support in the primary language
modality (either signed or spoken) as chosen by the family (Moeller et al., 2013) and may
include technological intervention (e.g., hearing aids), sign language consultation, or
spoken language intervention through speech-language pathology services. However, it is
not the case that families are committed to selecting one language modality. Rather,
given the publicly funded nature of the program, the IHP provides funding for families to
access services to support a primary language modality and families may pursue
additional, privately funded services if, for instance, they wish to raise their child in a
bimodal bilingual environment. Similarly, children in the IHP who are learning spoken
language may also be raised in homes with two spoken languages. In cases where
cochlear implantation is indicated, families access support through a collaboration with a separate publicly funded program and may not be followed by the IHP specifically. As a result, the present article focuses specifically on children who are hard of hearing (HH) and not children who are candidates for cochlear implantation. The IHP aligns its expectations closely with the recommendations put forth by the Canadian Infant Hearing Task Force and the JCIH. Currently, Ontario is one of six Canadian provinces/territories judged to be sufficiently meeting EHDI program standards (Canadian Infant Hearing Task Force, 2019).

Since 2009, spoken language outcome monitoring in the IHP has been conducted using the *Preschool Language Scale, 4th ed* (PLS-4; Zimmerman et al., 2006) every 6 months (JCIH 2007; 2013). Outcomes were to be tracked for all children for whom families selected spoken language as a primary language modality. This group can include children learning spoken language only or in conjunction with a signed language. Our research team was previously contracted by the IHP to evaluate outcomes using PLS-4 data from two birth cohorts in the program (Daub, 2016; Daub et al., 2017) and were therefore familiar with the previous process, as well as elements of data collection and reporting that were inconsistently implemented across the program. For example, less than 50% of the children in the birth cohorts analyzed did not have PLS-4 scores in the database, and PLS-4 scores were inconsistently scored across children (Daub, 2016). Because the nature of our involvement with the PLS-4 data was post-hoc, it was unclear whether data collection issues stemmed from issues with administration of the PLS-4, data entry/management errors, or errors in extraction from the data management system. The amount of data that were missing for undocumented reasons highlighted the
importance of improving upon the previous procedure to support program evaluation.

Around the same time that our team was involved in evaluating the outcome data from previous cohorts, the PLS-4 fell out of print in favor of the *Preschool Language Scale, 5th ed* (PLS-5; Zimmerman et al., 2011). As a result, the IHP sought to confirm that the PLS-5 would be an adequate replacement, and to evaluate and reconsider their procedure if necessary. At the same time, speech-language pathologists (SLPs) raised concerns about the appropriateness of the PLS-4/PLS-5 and questioned the rationale for its selection.

This article reports on a series of program evaluation and quality improvement projects we conducted to facilitate the IHP’s decision-making about a new spoken language outcome monitoring procedure. These projects began in 2014, and our initial recommendations were shared with the IHP in 2017. We begin by orienting the reader to the overall process we used to develop the procedure (see Figure 3-1). This includes identifying the IHP’s assessment purposes, developing a framework for assessing outcomes, and identifying tests to use in the framework. We then report on how we identified tests that appropriately fit within the framework, while also balancing needs at the level of both the program and the individual service providers and families.

### 3.1 Step 1: Identifying Assessment Purposes

#### 3.1.1 The IHP’s Assessment Purposes at the Program Level

Our main priority was to collect and maintain data within a provincial database that was appropriate for (a) evaluating the overall expressive and receptive spoken language outcomes of children in the IHP as a group to demonstrate the effectiveness of the IHP, (b) modeling children’s spoken language growth over time to identify
ages/stages of development where additional support might be needed, (c) identifying predictors of better, or worse, spoken language outcomes to support quality improvement initiatives, and (d) identifying whether there are differences in outcomes across regions of the province to support resource allocation. IHP management was also cognizant of the importance of clinician’s assessment purposes and minimizing the time and financial burden of spoken language outcome monitoring on service providers to the greatest extent possible. They were also interested in a procedure that could provide clinically useful data about individual children in addition to program-level evaluation.
3.1.2 The IHP’s Assessment Purposes for Individual Children and Families

At the level of the individual child and family, routine assessment of speech and language development should (a) identify children who are performing below age expectations and thus require speech-language development services, (b) allow profiling areas of relative strength and weakness in individual children, thus enabling clinicians to set goals and tailor interventions to meet individual needs at different stages of the child’s development, and (c) allow for evaluation of school readiness and anticipation of academic supports needed to ensure success upon school entry. Because children with permanent hearing loss have ongoing inconsistent access to auditory information, they are at greater risk for difficulties in certain areas of spoken language than others (Moeller, Tomblin, et al., 2007), even if they perform within age expectations on omnibus spoken language tests. Therefore, developing a procedure that is informative to intervention planning for individual children required an approach that probed more deeply than overall spoken language outcomes, specifically those domains of language that are (a) known to be at particular risk in children with permanent hearing loss and (b) predictive of future spoken language outcomes. For children with moderate to severe hearing loss, who are served by the IHP, there are certainly gaps in knowledge about development of specific spoken language domains (Moeller, Tomblin, et al., 2007), but some of the most vulnerable domains in children from birth to 6 years appear to be related to inconsistencies in auditory access, including:

1. Vocal development and canonical babbling in infancy (Moeller, Hoover, Putnam, Arbataitis, Bohenkamp, Peterson, Wood, et al., 2007; Moeller,


3. Morphosyntactic difficulty, which is suspected to stem from underlying concerns with articulation and phonology (Moeller, Tomblin, et al., 2007)

4. Phonological awareness in the preschool/kindergarten period (Moeller, Tomblin, et al., 2007)

3.1.3 Matching the Assessment Purpose with the Assessment Method

Achieving individual level purposes requires different assessment approaches and tests than achieving program level purposes. Individual level evaluation requires different tests measuring different vulnerabilities at different stages of development. Program level evaluation requires the same metric and the same or similar tests across programs and over time. To fulfill both of these sets of purposes, it became immediately apparent that there was no single test that would be sufficient.

As a result, we suggested a two-tiered outcome monitoring framework for the IHP: (a) monitoring overall receptive and expressive language development for program-level evaluation purposes using a single test, and (b) targeted individual monitoring of selected areas of speech/language vulnerability (see Figure 2). Although we recognize that concerns in any of these domains do not clearly begin or end at any age, we recommended limiting monitoring to selected areas of speech/language vulnerability using only one or two tests at any one of three developmental time points to minimize the
clinical burden of the process. This process was not intended to replace SLPs’ current practices of collecting the information they need to set goals and monitor progress for individual children on their caseload. Our next step was to identify which norm-referenced tests were best equipped to measure overall expressive and receptive spoken language and each of these domains.

3.2 Step 2: Selecting Tests for Outcome Monitoring

3.2.1 Step 2a) Scoping Review of Norm-Referenced Tests

The purpose of the scoping review was to identify which norm-referenced tests have been previously used in studies of children who are HH and the results obtained using each of these tests. In developing our recommendations, we sought to select amongst tests that to be sensitive enough to allow the IHP to detect group differences and change over time, should those differences or changes occur. Our expectation was that narrowing our consideration of norm-referenced assessments to only those that have been documented in the peer-reviewed literature would provide the IHP with benchmarks for spoken language outcomes, and some context to interpret their program’s results. We were cognizant that if we selected a set of tests that were not sensitive to group differences, or have not previously been used with children who are HH, then we ran the risk of overestimating the outcomes of children who are HH in the IHP. Inversely, if we selected tests that were very sensitive to the spoken language vulnerabilities of children who are HH, without appropriate research context to demonstrate that these results are reasonable, we ran the risk of underestimating children who are HH’s outcomes. Although age-have a documented history of use in the peer-reviewed literature as
preliminary evidence that the tests (a) have some ability to differentiate between children who are HH and children with typical hearing thresholds and (b) are sensitive to change over time. Although the original purpose of these studies was not to document test sensitivity to group differences per se, there is a dearth of norm-referenced tests designed specifically to capture the spoken language outcomes of children who are HH. Thus, our scoping review served as our closest approximation of whether a test was likely appropriate outcomes are appropriate goals for individual children who are HH, as a
group they have been demonstrated to statistically perform below their same-aged peers but within age-expectations (e.g., Ching et al., 2013). This is not to say that EHDI programs should not strive for spoken language outcomes on par with children who are typically hearing, per JCIH recommendations (2013). However, we did not want to over- or under-estimate the IHP’s impact based on artefacts of test selection.

Although EHDI intervention programs provide services to children and families electing to pursue spoken and signed language, and children who are (or are not) amplified with hearing aids or cochlear implants, our scoping review focused on articles reporting results of children who are HH who have been fitted with hearing aids and are learning a spoken language. In Ontario, cochlear implant candidacy represents a unique population who often receive services from a different publicly funded program and their outcomes are not routinely tracked by the IHP. We also restricted our review to outcomes measured in children who are HH from birth to 6 years of age to capture the language development of children who are HH in the program. Our initial review took place in 2016 across three databases (SCOPUS, CINAHL, and PubMed), but we conducted a more recent review across a modified set of databases for the purposes of this article to capture the most up-to-date publications. The results of this review were consistent with our prior review (Oram Cardy & Daub, 2017). Our review was guided by the following research questions:

1. Which tests have been used to measure spoken language in children who are HH who have been fitted with hearing aids between birth and 6 years?
2. Which tests have been used to compare children who are HH and children with typical hearing, or subgroups of children who are HH? Which tests have detected group differences?

3. Which tests have been used to measure change over time in children who are HH? Which tests have detected change over time?

3.2.1.1 Search Strategy

Five databases were searched in October 2018: CINAHL, Pubmed, EMBASE, ERIC, and PsycInfo. Search terms were developed with the assistance of a subject librarian (see Appendix 1 for an example search). The search was restricted to include only studies published between 1990 and 2018 to capture research completed during the time in which the evidence supporting universal newborn hearing screening and EHDI programs began to accumulate. Following the search, the titles, abstracts, and full texts of articles were screened for several criteria. First, the article must have been published in English. Second, the article needed to have measured spoken language using a commercially available, English, norm-referenced test. Third, the study was required to report outcome data for children who are HH who wore hearing aids separately from data for children who wore cochlear implants and needed to report data for, at a minimum, a subgroup of children between birth and 6 years, 11 months. Case studies of individual children where group data were not reported were also excluded.

Title, abstract, and full text screening from articles identified through the initial database search were completed by the first author and a trained research assistant to identify articles for full review. All eligibility disagreements were resolved through discussion. Title, abstract, and full text screening from articles identified through forward
and backward searching was completed by the first author using the same set of criteria previously described. This process was repeated until no new publications were identified.

The first author extracted from each eligible article: (a) the demographic characteristics of the study population; (b) the norm-referenced test(s), including test version, used; (c) whether group comparisons were made and the results of these comparisons; and (d) whether change over time was evaluated and the results of these evaluations. At this stage, studies were excluded if the norm-referenced test was out of print (i.e., studies using only the Reynell Language Developmental Scales; Reynell & Gruber, 1990). Older versions of tests were included if there is a more recent version available for purchase. Study quality was not evaluated as the purpose of our scoping review was to capture the breadth of tools used with children who are HH and the results found with them.

### 3.2.1.2 Scoping Review Results

We identified 12084 non-duplicate articles. Of those, 195 articles were retrieved after title and abstract screening. Finally, data were extracted from 36 articles (see Figure 3-3, and Supplemental Materials in Appendix 2 for the data extraction). From these 36 articles, 16 commercially available, norm-referenced tests across multiple versions were identified as having been previously used to measure spoken language outcomes in English-speaking children who are HH. Six of these tests were omnibus language measures, four were language or communication development subscales of broader developmental tests, three were measures of vocabulary, and three were measures of articulation and phonology.
For each test, the following was charted: the number of studies (out of 36) that used the test, whether any study used the test to make group comparisons (regardless of the results of the comparison), whether group differences were detected (out of the number of studies that used the test to evaluate group differences), whether any study used the test to measure change over time, and whether the test detected changes over time (out of the number of studies that used the test to evaluate group differences; see Table 3-1). Studies varied widely with respect to the ages of children included in the sample, the frequency with which they were assessed, the severity of hearing loss, characteristics of hearing aid amplification, and the demographics of comparison groups (see Supplemental Materials for further details). We identified a distinct lack of overlap in our studies in that no two studies evaluated the same outcomes in similar groups of children who are HH.
Table 3-1: Norm-referenced test use in research with children who are hard of hearing

<table>
<thead>
<tr>
<th></th>
<th># of studies that used the test for any purpose</th>
<th># of studies that used tests to compare groups</th>
<th># of studies that found group differences</th>
<th># of studies that measured change over time</th>
<th># of studies that detected change over time</th>
<th>Of studies using composite scores (n = 6), # of studies using test in composite score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Omnibus language tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS</td>
<td>15/36</td>
<td>8/15</td>
<td>5/8</td>
<td>3/15</td>
<td>3/3</td>
<td>4/6</td>
</tr>
<tr>
<td>MBCDI</td>
<td>9/36</td>
<td>7/9</td>
<td>4/7</td>
<td>2/9</td>
<td>1/2</td>
<td>0/6</td>
</tr>
<tr>
<td>CASL</td>
<td>4/36</td>
<td>2/4</td>
<td>1/2</td>
<td>0/4</td>
<td>n/a</td>
<td>2/6</td>
</tr>
<tr>
<td>PLAI</td>
<td>4/36</td>
<td>2/4</td>
<td>1/2</td>
<td>0/4</td>
<td>n/a</td>
<td>2/6</td>
</tr>
<tr>
<td>CELF</td>
<td>3/36</td>
<td>1/3</td>
<td>0/1</td>
<td>1/3</td>
<td>1/1</td>
<td>1/6</td>
</tr>
<tr>
<td>TACL</td>
<td>1/36</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
<td>1/1</td>
<td>0/6</td>
</tr>
<tr>
<td><strong>Language scales from developmental tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M)CDI</td>
<td>13/36</td>
<td>9/13</td>
<td>8/9</td>
<td>0/13</td>
<td>n/a</td>
<td>2/6</td>
</tr>
<tr>
<td>VABS</td>
<td>5/36</td>
<td>2/5</td>
<td>0/2</td>
<td>0/5</td>
<td>n/a</td>
<td>2/6</td>
</tr>
<tr>
<td>MSEL</td>
<td>2/36</td>
<td>1/2</td>
<td>1/1</td>
<td>0/2</td>
<td>n/a</td>
<td>1/6</td>
</tr>
<tr>
<td>WPPSI</td>
<td>1/36</td>
<td>0/1</td>
<td>n/a</td>
<td>0/1</td>
<td>n/a</td>
<td>1/6</td>
</tr>
<tr>
<td><strong>Vocabulary tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPVT</td>
<td>17/36</td>
<td>9/17</td>
<td>4/9</td>
<td>1/17</td>
<td>1/1</td>
<td>5/6</td>
</tr>
<tr>
<td>EVT</td>
<td>1/36</td>
<td>0/1</td>
<td>n/a</td>
<td>0/1</td>
<td>n/a</td>
<td>1/6</td>
</tr>
<tr>
<td>EOWPVT</td>
<td>2/36</td>
<td>1/2</td>
<td>1/1</td>
<td>1/2</td>
<td>1/1</td>
<td>0/6</td>
</tr>
<tr>
<td><strong>Articulation/phonology tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFTA</td>
<td>8/36</td>
<td>6/8</td>
<td>3/6</td>
<td>1/8</td>
<td>1/1</td>
<td>0/6</td>
</tr>
<tr>
<td>DEAP</td>
<td>6/36</td>
<td>2/6</td>
<td>2/2</td>
<td>0/6</td>
<td>n/a</td>
<td>4/6</td>
</tr>
<tr>
<td>KLPA</td>
<td>1/36</td>
<td>1/1</td>
<td>0/1</td>
<td>1/1</td>
<td>1/1</td>
<td>0/6</td>
</tr>
</tbody>
</table>

*Note. Six of the 36 reviewed studies used composite scores as an outcome measure. Multiple editions/versions of tests are combined. CASL = Comprehensive Assessment of Spoken Language; CELF = Comprehensive Evaluation of Language Fundamentals; DEAP = Diagnostic Evaluation of Articulation and Phonology; EOWPVT = Expressive One Word Vocabulary Test; EVT = Expressive Vocabulary Test; GFTA = Goldman-Fristoe Test of Articulation; KLPA = Khan-Lewis Phonological Analysis; MBCDI = MacArthur Bates Communicative Development Inventories; (M)CDI = (Minnesota) Child Development Inventory; MSEL = Mullen Scales of Early Learning; PLAI = Preschool Language Assessment Inventory; PLS = Preschool Language Scale; PPVT= Peabody Picture Vocabulary Test; TACL = Test of Auditory Comprehension of Language. VABS = Vineland Adaptive Behavior Scales.  
*Two studies only evaluated results descriptively.  
*bOne study only evaluated results descriptively.

Of the 36 studies identified, 30 used 16 different norm-referenced tests to compare spoken language outcomes to other children (i.e., children with typical hearing, with cochlear implants, or with different amplification technologies) or the test’s normative mean. Ten studies evaluated change over time using a variety of analyses (e.g., growth scale values, rates of language development, or linear regression). Six studies
evaluated spoken language outcomes using composite scores from multiple tests using factor analyses or multivariate analyses. Only 8 out of the 16 tests were used for both comparing spoken language outcomes to other groups of children and measuring change over time and none of the 8 tests consistently identified both differences between groups and change over time.

3.2.1.3 Scoping Review Implications

The scoping review provided 16 candidate tests for measuring each of the spoken language domains within the outcome monitoring process (see Figure 3-2). However, one of the tests (i.e., the Wechsler Preschool and Primary Scales of Intelligence; Wechsler, 2002) does not primarily measure spoken language, and largely measures domains that fall outside SLPs’ scope of practice in the province of Ontario. Therefore, it was excluded from future evaluations. Additionally, the Expressive Vocabulary Test (Williams, 2007) was used once in previous studies as a part of a composite score and was not used in studies making group comparisons or evaluating change over time. Given the lack of data about the Expressive Vocabulary Test’s performance on its own, we excluded it from future evaluations. Our next step was to examine the psychometric properties of each of the 14 candidate tests to determine which ones would be psychometrically appropriate to meet the IHP’s assessment purposes.

3.2.2 Step 2b) Critical Appraisal of Norm-Referenced Tests

After completing the initial 2016 scoping review, the most recent versions of the 14 tests, regardless of whether they were the versions used in studies included in the scoping review, were evaluated using the 2012 version of the Consensus Based Standards
for the Selection of Health Status Measurement Instruments (COSMIN; Mokkink et al., 2012) checklist. The COSMIN checklist was developed using an International Delphi study method where experts in fields related to measurement (e.g., epidemiology and statistics) iteratively responded to a series of questions about which measurement properties ought to be evaluated in test design (specifically Health-Related Patient Reported Outcomes, but with application to other tests) and the statistics that should be used to report them. Consensus (greater than 67% agreement) was reached on most major terms (with the exception of structural validity), definitions of each property, and on the taxonomy’s organization. From this taxonomy, the COSMIN team developed quality criteria for both the methodological quality of studies designed to collect data information about measurement properties, and the measurement properties themselves (Terwee, 2011). For the purposes of developing our recommendations, we focused our evaluation on the quality of the measurement properties reported in the examiner’s manual, but not the methodological quality of the studies designed to report the measurement properties, as it was quite likely that not all examiner’s manuals would report sufficient detail to adequately appraise the quality of the methods themselves.

3.2.2.1 Critical Appraisal Analysis

To appraise each test, we used a revised version of the COSMIN quality criteria in which we excluded four criteria that were included in the original checklist (criterion validity, cross-cultural validity, responsiveness, and measurement error). Although we agree that these criteria are important to consider, upon review it became clear that the statistics required to evaluate these criteria (e.g., differential item functioning analyses between multiple language versions) were very rarely evaluated in any of the included
tests, and evaluating these criteria would not support us in choosing a test amongst the 14 tests we identified. Therefore, each of the 14 tests were appraised with respect to the following: internal consistency, reliability, content validity, construct validity (hypothesis testing), and construct validity (structure). Each domain was assigned one of three ratings (positive, indeterminate, negative) according to the operationalizations of each criterion in the COSMIN checklist. For example, a test was rated as having positive evidence for structural validity if factors explained 50% or more of the variance, indeterminate if explained variance was not evaluated/discussed, or negative if factors explained 49% or less of the variance. For our purposes, we considered a test to have met reasonable criteria if they received a positive rating in at least 4 of the 5 of the categories.

3.2.2.2 Critical Appraisal Results

Only eight of the 14 of the tests met acceptable criteria in 4 of the 5 of the appraised COSMIN domains (see Table 3-2). Within each of the test categories (omnibus/language scale, vocabulary, phonology/articulation), at least one test met acceptable criteria in 4 of the 5 COSMIN domains. Most tests (12 of the 14) met acceptable criteria for reliability, and all tests reported at least one measure of reliability. Only one test reported weak evidence for validity domains, but most tests were missing validity information. Information about tests’ internal structure was the least frequently reported (only two of the 14 tests) in examiner’s manuals.

3.2.2.3 Critical Appraisal Implications

Based on our appraisal, we identified eight norm-referenced tests that were largely psychometrically acceptable to select for the spoken language outcome
Table 3-2: Critical appraisal of norm-referenced tests using COSMIN criteria

<table>
<thead>
<tr>
<th></th>
<th>Internal Consistency</th>
<th>Reliability</th>
<th>Content Validity</th>
<th>Hypothesis Testing</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Omnibus language tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS-5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>MBCDI-2</td>
<td>+/-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>CASL-2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>PLAI-2</td>
<td>?</td>
<td>+/-</td>
<td>?</td>
<td>+/-</td>
<td>+</td>
</tr>
<tr>
<td>CELF-P2</td>
<td>+/-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>TACL-4</td>
<td>+</td>
<td>?</td>
<td>?</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td><strong>Language scales from developmental tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDI</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>MSEL</td>
<td>?</td>
<td>+/-</td>
<td>?</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>VABS-3</td>
<td>+</td>
<td>+/-</td>
<td>?</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td><strong>Vocabulary tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPVT-4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>EOWPVT-4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td><strong>Articulation/phonology tests</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFTA-3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>DEAP</td>
<td>+</td>
<td>+/-</td>
<td>+</td>
<td>+/-</td>
<td>?</td>
</tr>
<tr>
<td>KLPA-3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
</tbody>
</table>

Note. Ratings included positive evidence (+), indeterminate (?), and negative evidence (−) in meeting COSMIN Criteria. +/- indicates that some, but not all, subtests meet acceptable criteria. Shaded tests received a positive rating in at least 4/5 of the categories. Preschool Language Scale (PLS-5; Zimmerman et al., 2011); MacArthur Bates Communicative Development Inventories (MBCDI-2; Fenson et al., 2007); Comprehensive Assessment of Spoken Language (CASL-2; Carrow-Woolfolk, 2017); Preschool Language Assessment Inventory (PLAI-2; Blank et al., 2003); Comprehensive Evaluation of Language Fundamentals (CELF-P2; Semel et al., 2004); Test of Auditory Comprehension of Language (TACL-4; Carrow-Woolfolk, 2014); Child Development Inventory (CDI; Ireton, 1992); Mullen Scales of Early Learning (MSEL; Mullen, 1995); Vineland Adaptive Behavior Scales (VABS-3; Sparrow et al., 2016); Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007); Expressive One Word Vocabulary Test (EOWPVT-4; Martin & Bronwell, 2011); Goldman-Fristoe Test of Articulation (GFTA-3; Goldman & Fristoe, 2015); Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd et al., 2006); Khan-Lewis Phonological Analysis (KLPA-3; Khan & Lewis, 2002).

monitoring process. There was not one test with clearly better measurement properties over the others. Our next step was to summarize the administration properties of each of these tests.

3.2.3 Step 2c) Consideration of Administration Properties

We considered various administration properties in summarizing the candidate tests including: the age ranges for which each test had normative data; whether the test
covered overall language abilities or subskills; the types of scores that could be calculated (e.g., percentile ranks and/or growth scale values), who was required to administer the test (clinician or caregiver), and the amount of time each test took to administer. Each of the eight acceptable tools had various administration properties that might make the test more, or less, attractive to individual EHDI programs (Table 3-3). For instance, the PLS-5, Clinical Evaluation of Language Fundamentals, Preschool, 2nd ed. (CELF-P2; Semel et al., 2004) and Comprehensive Assessment of Spoken Language, 2nd ed. (CASL-2; Carrow-Woolfolk, 2017) were all acceptable omnibus language measures, but the PLS-5 provides scores that support measuring change over time (i.e., growth scale values), the CELF-P2 supports profiling different domains of language, and the CASL-2 measures a broader range of language abilities and is appropriate at older ages than either the PLS-5 or CELF-P2. Therefore, consideration of these properties presented us with flexibility in which test(s) to propose. For the purpose of the IHP, tests like the PLS-5 had administration properties that would enable the IHP to achieve more of their outcome monitoring purposes. Specifically, the PLS-5 reported normative data for all age ranges served by the program and also reported growth scale values, which would enrich program level evaluation of growth over time. However, other tests had other relative advantages over the PLS-5. For instance, the MacArthur-Bates Communicative Development Inventories, 2nd ed. (MBCDI-2; Fenson et al., 2007) could be completed by parents without SLPs’ support, and the CELF-P2 supported profiling. Our next step was to triangulate the administrative properties and relative advantage of each test with the evidence for the quality of each test to develop a set of options. We then shared these
initial recommendations with the IHP and a panel of expert SLPs who had volunteered their time to provide feedback on the clinical feasibility of our recommendations.

Table 3-3: Administration properties for currently available versions of psychometrically suitable for norm-referenced tests

<table>
<thead>
<tr>
<th>Age range</th>
<th>Language Areas</th>
<th>Scores Available</th>
<th>Examiner/Respondent</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
<td>Subskills</td>
<td>SS</td>
<td></td>
</tr>
<tr>
<td>PLS-5</td>
<td>0-7 years</td>
<td>✓</td>
<td>✓</td>
<td>Clinician</td>
</tr>
<tr>
<td>MBCDI-2</td>
<td>8-18, 16-30,</td>
<td>✓</td>
<td>✓</td>
<td>Caregiver</td>
</tr>
<tr>
<td></td>
<td>30-37 months</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASL-2</td>
<td>3-6 years</td>
<td>✓</td>
<td>✓</td>
<td>Clinician</td>
</tr>
<tr>
<td>CELF-P2</td>
<td>3-6 years</td>
<td>✓</td>
<td>✓</td>
<td>Clinician</td>
</tr>
<tr>
<td>PPVT-4</td>
<td>2.6-90 years</td>
<td>✓</td>
<td>✓</td>
<td>Clinician</td>
</tr>
<tr>
<td>GFTA-3</td>
<td>2-21 years</td>
<td>✓</td>
<td>✓</td>
<td>Clinician</td>
</tr>
<tr>
<td>DEAP</td>
<td>3-8 years</td>
<td>✓</td>
<td>✓</td>
<td>Clinician</td>
</tr>
<tr>
<td>KLPA-3</td>
<td>8-21 years</td>
<td>✓</td>
<td>✓</td>
<td>Clinician</td>
</tr>
</tbody>
</table>

Note. AE = age equivalent; GSV = gross scale value; PR = percentile rank; SS = standard score.

Preschool Language Scale (PLS-5; Zimmerman et al., 2011); MacArthur Bates Communicative Development Inventories (MBCDI-2; Fenson et al., 2007); Comprehensive Assessment of Spoken Language (CASL-2; Carrow-Woolfolk, 2017); Comprehensive Evaluation of Language Fundamentals (CELF-P2; Semel et al., 2004); Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007); Goldman-Fristoe Test of Articulation (GFTA-3; Goldman & Fristoe, 2015); Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd et al., 2006); Khan-Lewis Phonological Analysis (KLPA-3; Khan & Lewis, 2002).

3.3 Step 3: Integrating the Evidence into Recommendations

3.3.1 Recommendations for Overall Spoken Language Outcome Monitoring

In accordance with JCIH recommendations, we proposed that all children in the IHP be tested with a standardized measure that compares their spoken language development to that of same-aged children with typical hearing every 6 months during the first 3 years of life, and every year thereafter. Triangulation of the evidence from our scoping review, critical appraisal, and summary of administration properties indicated that the following three measures had the strongest evidence supporting their selection as a measure of overall language abilities: PLS-5, MBCDI-2, and CELF-P2. Both the PLS-5
and CELF-P2 offer the additional advantages of having diagnostic accuracy information with cut-point scores and growth scale values. The PLS-5 covers the full 0 to 6 year age range serviced by the IHP, while the CELF-P2 covers 3 to 6 years, and the MBCDI includes three separate forms that cover 8 to 18 months (MBCDI Words and Gestures), 16 to 30 months (MBCDI Words and Sentences), and 30 to 37 months (MBCDI III). Therefore, the most parsimonious approach would be to use the PLS-5 across the entire age span of the program. However, we have encountered SLPs and scientific experts in the field of permanent childhood hearing loss (e.g., Dr. Mary Pat Moeller, personal communication) who have expressed concerns about the sensitivity of the PLS-5 in the first two years of life. These concerns are consistent with the diagnostic accuracy data reported in the examiner’s manual (Zimmerman et al., 2011). That is, the PLS-5’s diagnostic accuracy does not meet acceptable criterion ($\geq 0.80$; Plante & Vance, 1994) for detecting language delays in children under 2 years for any cut-score. Therefore, although using the PLS-5 would allow the IHP to evaluate whether children were making significant progress over time, SLPs would be unable to accurately determine whether children were obtaining age-appropriate outcomes and the PLS-5 posed greater clinical burden (i.e., longer administration time) than other candidate tests.

An alternative option could be to use the three separate forms of the MBCDI-2 in the first three years of life and the CELF-P2 thereafter. However, because the subtests and scores on the three MBCDI-2 forms are different, this would prohibit future analysis of developmental growth over time, which “can only be analyzed if the child is assessed with at least some instruments that can be repeated throughout the target age range” (JCIH, 2013, p. e1334). An additional concern is that only the MBCDI Words and
Gestures form includes evaluation of both receptive and expressive language (along with gestures); the remaining MBCDI-2 forms only assess expressive language.

A third option included using the MBCDI-2 Words and Gestures form until 18 months of age, and the PLS-5 thereafter. This would provide scores on the same measure (the MBCDI-2) for the first two testing sessions at the 6-month testing interval, and then PLS-5 scores for all 6-month and 12-month testing intervals beyond 18 months. Under this option, the program would be able to make direct comparisons of growth across all time points except for the one point of transition between the MBCDI-2 and PLS-5 around 18 to 24 months. We felt that this was a reasonable compromise to have a more clinically accepted tool in the earliest years of development, and thus this third option formed the basis for our final recommendation.

3.3.2 Recommendations for Individual Vulnerability Testing

Our scoping review and critical appraisal identified norm-referenced tests that have been used with children who are HH and that measure areas that are particularly vulnerable for them. Based on the results of our scoping review and critical appraisal, we recommended a two-pronged approach to assessment for the purposes of supporting individual child/family needs. We recommended that SLPs include assessment of key vulnerabilities associated with the child’s particular age/stage of development (see Figure 4) alongside of their administration of the program-level test of overall language abilities. To reduce the time associated with assessment, and to prevent children from being assessed with more than two norm-referenced tests at a single session, we recommended assessing one area of key vulnerability at each age, even though the ages at which different skills (e.g., articulation and phonology) can be assessed may overlap with other
key vulnerabilities. Additionally, in our scoping review we were unable to identify any commercially available test of early vocal development, although some articles (e.g., Ambrose et al., 2014) report on experimental tests that are currently in development. In this regard, we were unable to recommend a specific test for the IHP to use for monitoring early vocal development. In short, we recommended that the IHP provide a set of recommended tests from which SLPs are advised to select. This would support consistency across regions and ensure that only those tests with the strongest evidence are used to assess these key vulnerability areas.

3.4 Consultation with Stakeholders

We summarized the overall process (program level monitoring and individual vulnerability testing) as well as the three options for overall outcome monitoring and our recommendations for individual vulnerability testing (described above), in a formal written report (Oram Cardy & Daub, 2017). This report was shared with IHP audiological policy development, IHP government leaders, and a team of SLPs who formed an advisory panel. All parties provided written feedback on the report and discussed the recommendations at length through teleconference meetings. Following the revisions to the recommendations, all parties reached agreement on a final procedure (see Figure 3-4). This procedure included program-level outcome monitoring and individual vulnerability testing. Following final discussion via teleconference, the managerial team ultimately adopted the final spoken language outcome monitoring procedure for implementation in the IHP.
3.5 Discussion

The present article describes our process for developing a set of spoken language outcome monitoring recommendations to support a Canadian EHDI program, the Ontario Figure 3-4: Final recommendation

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Program Monitoring</th>
<th>Individual Vulnerability Testing</th>
<th>Emergent literacy/Phonological awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5-1</td>
<td>MBCDI-2 Words &amp; Gestures* <em>(Scores for: Words Understood, Words Produced, Phrases Understood, and Gestures Produced)</em></td>
<td>Vocalization/Babbling/Articulation/Phonology</td>
<td>MBCDI-2 Words &amp; Gestures* <em>(Scores for: Words Understood, Words Produced, Phrases Understood, and Gestures Produced)</em></td>
</tr>
<tr>
<td>1-1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5-2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5-3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>PLS-5 <em>(Scores for: Auditory Comprehension &amp; Expressive Communication)</em></td>
<td>Vocal development tests require further evaluation</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. CASL-2 = Comprehensive Assessment of Spoken Language; CELF-P2 = Comprehensive Evaluation of Language Fundamentals; EOWPVT = Expressive One Word Vocabulary Test; GFTA-3 = Goldman-Fristoe Test of Articulation; MBCDI-2 = MacArthur Bates Communicative Development Inventories; PLS-5 = Preschool Language Scale.
IHP, in fulfilling best practice recommendations. To date, there has been limited guidance in the literature on (a) the best way to approach the development of a spoken language outcome monitoring process or (b) how to accomplish all of the facets of spoken language outcome monitoring in a way that provides statistically appropriate evidence, is implementable across entire EHDI programs, and meets the competing needs of different stakeholders. Our expectation is that documenting our steps in this process and the recommendations that resulted will not only provide a general framework and example for other EHDI programs, but also highlight the previously undiscussed challenges of designing such a procedure.

Our process was grounded in the initial JCIH (2007, 2013) recommendations for spoken language as well as consideration of the International Consensus work on best practice principles (Moeller et al., 2013). From this foundation, we considered the purposes of spoken language outcome monitoring from the perspective of various IHP stakeholders to clarify the assessment purposes our process would need to fulfill. Using these purposes, we conducted a scoping review to identify a set of candidate norm-referenced tests that have been previously used to fulfill these assessment purposes and appraised the psychometric quality of the most recent versions of these tests. We then considered the administration properties of the tests that we rated as psychometrically acceptable and integrated all sources of evidence with our originally described assessment purposes. This allowed us to develop a set of recommendations to share with IHP stakeholders, who ultimately decided to adopt them. We expect that our work will be of interest to other EHDI programs and service providers who work with children who are DHH by documenting our process in developing our recommendations, the
recommendations themselves, and the final procedure adopted by the IHP. Our results highlight the unique challenges faced when trying to develop a process for spoken language outcome monitoring, guide future research designed to refine the development process, and contribute to a body of literature that provides guidance for EHDI programs looking to fulfill best practice recommendations.

Our next step is to design implementation materials and conduct pilot projects to evaluate the new procedures for both overall spoken language monitoring and individual vulnerability monitoring. These pilot projects are intended to identify barriers and facilitators to implementing the new recommendations in clinical practice, and to allow us to refine our process into one that is most sustainable and clinically feasible before program-wide launch. We anticipate that the results of these pilot projects will similarly support discussions of spoken language outcome monitoring in EHDI programs and highlight the inherent complexity in accomplishing these goals.

We do not intend to assert that our process or final recommendations are a gold standard for spoken language outcome monitoring and should be adopted by other EHDI programs. Rather, we believe that our work uniquely highlights the challenges in accomplishing spoken language outcome monitoring and may be a valuable foundation for EHDI programs looking to develop, or refine, their spoken language outcome monitoring procedures. Our projects were developed through the lens of the Ontario IHP, and other EHDI programs might have different priorities for spoken language outcome monitoring, amongst other needs. In our case, the IHP sought a process that would allow them to use the data to evaluate whether children across the province are making progress in their spoken language over time, whether they are meeting age-appropriate
expectations by the time they are discharged from the program, and whether they have the spoken language skills they need at discharge to be prepared for school. Necessarily, fulfilling these purposes required the use of multiple tests that are sensitive to multiple domains of language, and that were norm-referenced to establish whether a child was performing within or below age-expectations.

An additional priority was selecting norm-referenced tests from those that have been previously used in research with children who are HH to contextualize the outcomes in the IHP with the peer-reviewed literature. The Ontario IHP is publicly funded and managed under a larger provincial division also responsible for the allocation of resources across multiple programs from a single budget. We were wary of selecting norm-referenced tests without a documented history of use in the literature because it has been demonstrated that children who are HH often score within age-expectations (and close to the test’s normative mean of a standard score of 100), but statistically lower than matched groups of children with typical hearing (e.g., Tomblin et al., 2015). In this case, using a standard score cut-off recommended by a norm-referenced test was not sufficient to describe program outcomes. We were aware that spoken language outcome data could be used by policy makers to make funding decisions and that there was a risk of misinterpreting program level outcomes as being insufficient to continue funding. We were also aware that EHDI programs are precariously positioned in Canada: many EHDI programs are in development, and some have seen declines in support from previous years (Canadian Infant Hearing Task Force, 2014; 2019). In the Canadian context, statistically sound outcome data from one EHDI program has the potential to provide evidence to influence other provincial or national funding priorities. Therefore, it was
critical to develop a process that we could connect to the peer-reviewed literature to evaluate whether the IHP was performing on par with documented outcomes in other EHDI programs.

Even within the context of the Ontario IHP, our recommendations remain limited in a number of respects. Canada has two official languages (English and French) and many regions in the province are densely populated, multicultural areas where residents speak languages other than these. We focused our reviews and recommendations on measuring outcomes for children who are HH from English speaking families, in part, due to a dearth of norm-referenced tests that have been validated in other languages to include in our scoping review and critical appraisal. Certainly, many (but not all, i.e., the MBCDI-2) of the tests we selected for our current recommendations have not been normed in French, even if there are translated versions (i.e., the PLS-5). To fulfil clinical assessment needs, we have advised SLPs to continue using the tools they typically would for children for whom English is not a primary language, although their outcomes will not be able to be evaluated at the program-level in the provincial database. This raises concerns about equitable service provision—regardless of the language their child is learning, families deserve to know whether their child is progressing as expected in response to intervention. Solutions and next steps, such as collecting local normative data on translated versions, are under discussion. Until norm-referenced assessments for these groups of children exist, EHDI programs will need to identify other creative solutions to evaluate spoken language outcomes and rely on less formal assessments. Our general framework could be modified to support identifying informal assessments or interview
tools, although a different process for critically appraising the approaches would be needed.

It is likely that there are other important considerations requiring attention in other EHDI programs that we did not account for in our process for the Ontario IHP. For example, EHDI programs in which outcome data are not likely to be used to support funding decisions may feel comfortable considering the use of norm-referenced tests without a history of previous peer-reviewed use. Additionally, our process did not consider the spoken language outcomes of children with cochlear implants because many are served by a different program in the province of Ontario, but other EHDI programs may wish to do so. Furthermore, our process did not attend to the sensitivity and specificity cut-off scores for language impairment on the tests we evaluated because there is no mandate in Ontario for children to perform below a certain threshold (e.g., -2 SD below the mean) to be considered eligible for receiving SLP services outside of EHDI programming. This is certainly the case in some American state education departments (Spaulding et al., 2012), thus, EHDI programs located in regions with similar requirements will need to additionally consider whether candidate tests are adequately sensitive/specific at the cut-off scores required to receive services.

Despite these limitations, our experience has highlighted major challenges in fulfilling spoken language outcome monitoring worthy of further consideration by the field. There is certainly more room for discussion about which assessment considerations ought to be prioritized in developing spoken language outcome monitoring procedures, the role of norm-referenced tests versus other sources of assessment information (e.g., criterion referenced testing for goal setting), and ways to ensure equity in how these
sources of information are collected and used across programs. First, outcomes from two norm-referenced tests are not directly comparable and the operationalization of “within age-expectations” is entirely dependent on the statistical properties of the norm-referenced test in question. Although the JCIH recommends that children who are HH should score within -1 SD of the mean or higher on norm-referenced tests (2013), this recommendation does not acknowledge the unique sensitivity and specificity of individual tests at individual scores (Spaulding et al., 2006). For example, both the PLS-5 and the CELF-P2 have the greatest diagnostic accuracy at -1 SD (Zimmerman et al., 2011; Semel et al., 2004), but the GFTA-3 maximizes diagnostic accuracy at -1.5 SD (Goldman & Fristoe, 2015). As such, children with typical hearing thresholds and typical language development can be expected to score between -1.49 and -1 SDs below the mean on the GFTA-3. If stakeholders apply the -1 SD cut-off as the expectation on tests that are less accurate at -1 SD, they may be inadvertently holding children who are HH to a higher standard than their peers with typically developing hearing. In other words, defining age-appropriate outcomes for individual children, and appropriate outcomes for children who are HH as a group, is confounded with the psychometric properties of norm-referenced tests (Spaulding et al., 2006). These confounds pose significant challenges to stakeholders looking to interpret their population level outcome data. A program that elects to use the PLS-5 to measure outcomes might appear to have better outcomes (i.e., within -1 SD of the mean) than a program that elects to use a test with a -1.5 SD cut-off, even though the children in both programs might be performing within age-expectations. Therefore, procedures for measuring outcomes must consider the
unique psychometric properties of the tests they are using or risk generating data that suggests their program is failing to meet JCIH benchmarks.

These concerns with defining age-appropriate outcomes and interpreting results are compounded when we consider applying spoken language outcome monitoring to different groups of children, including those 20% to 40% of children who are HH who have additional diagnoses, some of which (e.g., autism, cerebral palsy, and developmental delay) may further impact language development (Cupples, Ching, Crowe, Day, et al., 2014). Future work could extend the methods used here to identify studies examining language outcomes in children with an additional diagnosis, with and without hearing loss. This would provide context to any program looking to report on the results of children who are HH with additional disabilities.

A second challenge with accomplishing spoken language outcome monitoring pertains to the clinical feasibility of accomplishing all necessary assessment purposes. Many norm-referenced tests are not developed to serve multiple assessment purposes, and their use is best restricted to interpreting whether a child is, or is not, within age-expectations. This creates challenges for accomplishing the diverse purposes that spoken language outcome monitoring is intended to fulfill (e.g., treatment planning and evaluating EHDI programs broadly). Some of these purposes can certainly be accomplished through other forms of assessment (e.g., criterion referenced assessment, language sample analysis), and neither we, nor the JCIH (2013), argue that norm-referenced assessments should be the only component of a spoken-language outcome monitoring battery. Certainly, SLPs will need to rely on other sources of information to develop their therapy plans. However, the addition of a standard norm-referenced process
to fulfill program-level evaluation goals adds lengthy tasks to SLPs’ assessment time and it is unknown whether it is feasible for SLPs to collect, interpret and integrate all of the necessary sources of information needed to fulfill spoken language outcome monitoring recommendations. It is widely accepted that whether research evidence or new recommendations will be successfully used in clinical practice is influenced by numerous factors within the clinical context (e.g., Dobrow et al., 2004; Graham et al., 2006) such as time, caseload, and clinician factors (e.g., beliefs, knowledge, skills) above and beyond the quality of the research evidence or recommendation itself. Accomplishing spoken language outcome monitoring in EHDI programs is complicated not only by limited evidence to guide development of procedures, but also by a lack of evidence to support implementation of these procedures. To our knowledge, there is only one peer-reviewed paper, published by our research group (Cunningham et al., 2019) that has evaluated SLPs’ perceptions of the barriers to implementing spoken language outcome monitoring in an EHDI program. In Cunningham’s investigation time for additional testing was a primary concern. Additional work is needed to evaluate the feasibility of our recommendations specifically, and spoken language outcome monitoring broadly, as well as to develop implementation interventions that result in effective, sustained uptake of spoken language outcome monitoring procedures.

3.6 Conclusions

Guidance for how to best implement spoken language outcome monitoring recommendations (JCIH 2007; 2013) is lacking, and EHDI programs face significant barriers to developing procedures that fulfill best-practice recommendations. The present paper describes a series of projects, conducted as part of program evaluation and quality
improvement for the Ontario IHP, to develop a spoken language outcome monitoring procedure using a scoping review and critical appraisal of candidate norm-referenced tests. We expect that the process we used, the recommendations we developed, and the challenges we encountered, will be informative to other EHDI programs looking to develop their own procedures. Final recommendations included developing a two-tiered assessment battery measuring overall spoken language outcomes and key areas of spoken language vulnerability. Future work evaluating the appropriateness of these recommendations, whether the data collected is sufficient to fulfill our intended purposes, the feasibility of our recommendations and ways to implement them into clinical practice are needed.
References


https://doi.org/10.1044/2015_JSLHR-L-14-0134


https://doi.org/10.1097/AUD.0000000000000213

https://doi.org/10.1542/peds.106.3.e43


Chapter 4

Usability and feasibility of a Spoken Language Outcome Monitoring Procedure in a Canadian Early Hearing Detection & Intervention Program: Results of a 12-month Pilot

Early Hearing Detection and Intervention (EHDI) program best practice recommendations 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 of Infant Hearing, 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; Joint Committee on Infant Hearing, 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 previously used the *Preschool Language Scale, 4th edition* (Zimmerman, Steiner & Pond, 2002) to monitor spoken language outcomes. When the *Preschool Language Scale, 4th edition* fell out of print, the IHP contracted the authors to support developing 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 have previously been used in the peer-reviewed literature to evaluate children’s performance. Based on the results of a scoping review, critical appraisal, and consultation with IHP managers and speech-language pathologists (SLPs), a two-tiered assessment approach was recommended. In Tier 1, it was recommended that SLPs measure spoken language every six months from birth to 3;0, and annually thereafter (Joint Committee of Infant Hearing, 2007; 2013). Between birth and 1;6, 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;7 to 6;0, the *Preschool Language Scale, 5th edition* (PLS-5; Zimmerman, Steiner & Pond, 2011) was recommended. The PLS-5 was selected based on its suitability for children within IHP age eligibility (up to 6;0), 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 4-1 & 4-2). 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.

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 4-1 & 4-2). This tier was recommended as an improvement to the existing standard of care whereby children were discharged from SLP 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 4-2). It was recommended that SLPs track key vulnerabilities at the same intervals as overall
Figure 4-1: Overall outcome monitoring process, Daub & Oram Cardy (2021)

**Tier 1: Overall Language Assessment (Birth – Program Discharge):**

Goal: Assess overall receptive and expressive language development with an omnibus tool for program outcome evaluation

**Tier 2: Key Vulnerability Monitoring (Birth – Program Discharge):**

Goal: Measure specific domains of spoken language known to be at risk in children who are hard of hearing

- **Key Vulnerabilities in Infants**
  - Goal: Monitor vocalization and babble development for predictors of speech-language delay in individual children

- **Key Vulnerabilities in Toddlers**
  - Goal: Monitor speech-language development with attention to speech and first words to identify individual children in need of additional speech-language development support

- **Key Vulnerabilities in Preschoolers/Kindergarteners**
  - Goal: Monitor morpho-syntax and emergent literacy/phonological awareness development to tailor intervention and identify additional supports needed for school success in individual children
Figure 4-2: Tests used in outcome monitoring process, Daub & Oram Cardy (2021)

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Program Monitoring</th>
<th>Individual Vulnerability Testing</th>
<th>Emergent literacy/Phonological awareness</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5-1</td>
<td>MBCDI-2 Words &amp; Gestures*&lt;br&gt;(Scores for: Words Understood, Words Produced, Phrases Understood, and Gestures Produced)</td>
<td>Vocalization/Babbling/&lt;br&gt;Articulation/Phonology</td>
<td>(MBCDI-2 Words &amp; Gestures)</td>
</tr>
<tr>
<td>1-1.5</td>
<td>GFTA-3 (Scores for Sounds-in-Words)</td>
<td>Vocal development tests require further evaluation</td>
<td></td>
</tr>
<tr>
<td>1.5-2</td>
<td>MBCDI-2 Words &amp; Sentences or EOWPVT-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5-3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>CELF-P2 (Scores for Word Structure) or CASL-2 (Scores for Grammatical Morphemes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>CELF-P2 (Scores for Pre-literacy Rating Scale) or CELF-P2 (Scores for Phonological Awareness Subtest)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. CASL-2 = Comprehensive Assessment of Spoken Language; CELF-P2 = Comprehensive Evaluation of Language Fundamentals; EOWPVT = Expressive One Word Vocabulary Test; GFTA-3 = Goldman-Fristoe Test of Articulation; MBCDI-2 = MacArthur Bates Communicative Development Inventories; PLS-5 = Preschool Language Scale.
spoken language (every six months from birth to 3;0 and annually afterwards). 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. This paper reports data from two pilot studies that were initiated to evaluate the usability and feasibility of both tiers prior to program-wide implementation. 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. Specifically, we addressed the following questions for each tier:

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?

4.1 Pilot Study 1: Tier 1 Program-level Outcome Monitoring

4.1.1 Method

4.1.1.1 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.

4.1.1.2 Procedure

Prior to implementing the pilot program, participating SLPs (N=56) from eleven 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 extracted data for analysis and checked all test scores for typographical or scoring errors. Unique identification numbers were used to extract additional clinical information (i.e., child’s sex, audiological variables) from the IHP database. This database is 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. 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.

To identify barriers to implementation and modifications 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.

4.1.1.3 Participants: Children assessed in the pilot

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. 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. 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 a subset of 134 children who had bilateral sensorineural or mixed hearing losses. Although the purpose of the Program-level 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 Program-level outcome monitoring procedure was sensitive to known predictors of spoken language outcomes, we elected to focus our analyses on the groups of children for whom there is the most peer-reviewed data to contextualize our outcomes - children with bilateral sensorineural or mixed hearing losses.

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 4-1) and 34 had data for two assessments (see Table 4-2). Twenty-eight children had data for at least one assessment with the MBCDI-2 (see Table 4-3) and nine had data for two assessments (see Table 4-4). 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 Appendices 2-6 for a comparison of aided SII to 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
Table 4-1: Demographics of children with data for one PLS-5 assessment

<table>
<thead>
<tr>
<th>Variable</th>
<th>Children without Additional Factors (N=75)</th>
<th>Children with Additional Factors (N=41)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M ) (range)</td>
<td>( SD )</td>
</tr>
<tr>
<td>Age (months)</td>
<td>38.3 (19 – 71)</td>
<td>7</td>
</tr>
<tr>
<td>BEPTA (dB HL)</td>
<td>53.2 (17.5 – 107.5)</td>
<td>23.2</td>
</tr>
<tr>
<td>Better Ear SII (Conversational Speech)</td>
<td>72.5 (5 – 95)</td>
<td>22.59</td>
</tr>
<tr>
<td>Better Ear SII (Quiet Speech)</td>
<td>64.29 (2 – 97)</td>
<td>24.83</td>
</tr>
<tr>
<td>Expressive Communication (Standard Score)</td>
<td>100.92 (50 – 150)</td>
<td>20.5</td>
</tr>
<tr>
<td>Auditory Comprehension (Standard Score)</td>
<td>98.96 (50 – 137)</td>
<td>19.81</td>
</tr>
</tbody>
</table>

Note: BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss

Table 4-2: Demographics of children with data for two PLS-5 assessments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Children without Additional Factors (N=24)</th>
<th>Children with Additional Factors (N=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M ) (range)</td>
<td>( SD )</td>
</tr>
<tr>
<td>Age at first PLS-5 (months)</td>
<td>26.96 (19 – 38)</td>
<td>6.17</td>
</tr>
<tr>
<td>Age at second PLS-5 (months)</td>
<td>34.76 (24 – 48)</td>
<td>7.04</td>
</tr>
<tr>
<td>BEPTA (dB HL)</td>
<td>55.55 (20 – 107.5)</td>
<td>25.04</td>
</tr>
<tr>
<td>Better Ear SII (Conversational Speech)</td>
<td>69.61 (5 – 95)</td>
<td>26.13120</td>
</tr>
<tr>
<td>Better Ear SII (Quiet Speech)</td>
<td>61.11 (13 – 97)</td>
<td>26.70732</td>
</tr>
<tr>
<td>First Expressive Communication (Standard Score)</td>
<td>103 (73 – 123)</td>
<td>14.07</td>
</tr>
<tr>
<td>Second Expressive Communication (Standard Score)</td>
<td>101.9 (74 – 122)</td>
<td>14.95</td>
</tr>
<tr>
<td>First Expressive Communication (Growth Scale Value)</td>
<td>382.25 (297 – 448)</td>
<td>36.99</td>
</tr>
<tr>
<td>Second Expressive Communication (Growth Scale Value)</td>
<td>412.5 (314 – 507)</td>
<td>43.79798</td>
</tr>
<tr>
<td>First Auditory Comprehension (Standard Score)</td>
<td>104.35 (81 – 127)</td>
<td>13.94</td>
</tr>
<tr>
<td>Second Auditory Comprehension (Standard Score)</td>
<td>103 (65 – 123)</td>
<td>14.72</td>
</tr>
<tr>
<td>First Auditory Comprehension (Growth Scale Value)</td>
<td>394.45 (324 – 450)</td>
<td>34.27</td>
</tr>
<tr>
<td>Second Auditory Comprehension (Growth Scale Value)</td>
<td>426.74 (352 – 504)</td>
<td>36.1</td>
</tr>
</tbody>
</table>
Comprehension (Growth Scale Value)

*Note: BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss*

### Table 4-3: Demographics of children with data for one MBCDI assessment

<table>
<thead>
<tr>
<th>Variable</th>
<th>Children without Additional Factors (N=19)</th>
<th>Children with Additional Factors (N=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M ) (range)</td>
<td>( SD )</td>
</tr>
<tr>
<td>Age (months)</td>
<td>12.37 (8 – 18)</td>
<td>3.14</td>
</tr>
<tr>
<td>BEPTA (dB HL)</td>
<td>56.23 (31.25 – 95)</td>
<td>19.77</td>
</tr>
<tr>
<td>Better Ear SII</td>
<td>72 (21 – 91)</td>
<td>22.77</td>
</tr>
<tr>
<td>(Conversational Speech)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better Ear SII (Quiet Speech)</td>
<td>64.17 (6 – 88)</td>
<td>27.31</td>
</tr>
<tr>
<td>Phrases Understood (Percentile Rank)</td>
<td>37.5 (10 – 75)</td>
<td>19.8</td>
</tr>
<tr>
<td>Words Produced (Percentile Rank)</td>
<td>32.78 (&lt;5 – 85)</td>
<td>29.67</td>
</tr>
<tr>
<td>Words Understood (Percentile Rank)</td>
<td>42 (10 – 99)</td>
<td>25.85</td>
</tr>
<tr>
<td>Gestures (Percentile Rank)</td>
<td>39.67 (&lt;5 – 80)</td>
<td>22.61</td>
</tr>
</tbody>
</table>

*Note: BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss*

### Table 4-4: Demographics of children with data for two MBCDI assessments

<table>
<thead>
<tr>
<th>Variable</th>
<th>Children without Additional Factors (N=5)</th>
<th>Children with Additional Factors (N=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M ) (range)</td>
<td>( SD )</td>
</tr>
<tr>
<td>Age at first MBCDI (months)</td>
<td>10.5 (8 – 14)</td>
<td>2.65</td>
</tr>
<tr>
<td>Age at second MBCDI (months)</td>
<td>15.25 (14 – 17)</td>
<td>1.26</td>
</tr>
<tr>
<td>BEPTA (dB HL)</td>
<td>54.5 (31.25 – 90)</td>
<td>25.6</td>
</tr>
<tr>
<td>Better Ear SII (Conversational Speech)</td>
<td>78 (71 – 85)</td>
<td>7</td>
</tr>
<tr>
<td>Better Ear SII (Quiet Speech)</td>
<td>75.67 (64 – 82)</td>
<td>10.15</td>
</tr>
<tr>
<td>First Phrases Understood (Percentile Rank)</td>
<td>28.75 (14 – 45)</td>
<td>13.77</td>
</tr>
<tr>
<td>Second Phrases Understood (Percentile Rank)</td>
<td>28.75 (15 – 40)</td>
<td>11.09</td>
</tr>
<tr>
<td>First Words Produced (Percentile Rank)</td>
<td>50 (5 – 80)</td>
<td>31.88</td>
</tr>
<tr>
<td>Second Words Produced (Percentile Rank)</td>
<td>30 (25 – 40)</td>
<td>7.01</td>
</tr>
<tr>
<td>First Words Understood (Percentile Rank)</td>
<td>43.75 (20 – 55)</td>
<td>16.01</td>
</tr>
<tr>
<td>Second Words Understood (Percentile Rank)</td>
<td>30 (10 – 50)</td>
<td>16.83</td>
</tr>
<tr>
<td>First Gestures (Percentile Rank)</td>
<td>36.25 (5 – 60)</td>
<td>22.23</td>
</tr>
<tr>
<td>Second Gestures (Percentile Rank)</td>
<td>37.5 (15 – 50)</td>
<td>15.55</td>
</tr>
</tbody>
</table>

*Note: BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss*
4.1.2 Analyses: Data usability

There were two primary analytic purposes of the Tier 1 Program-level 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. These predictors included the severity of hearing loss and the presence/absence of additional factors influencing performance. Additional factors were broadly defined as any factor that an 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. 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 scores were consistently entered 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 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.

Once corrected, PLS-5 and MBCDI-2 scores were entered into a direct entry linear regression model, using pure tone average hearing thresholds and the dichotomous coding of the presence/absence of additional factors that SLPs believed may have influenced a child’s performance as independent variables. The influence of severity of hearing loss was evaluated using a separate linear regression model. For children with
bilateral hearing loss, the Better Ear Pure Tone Average (BEPTA) was entered as the predictor of growth. 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). will attempt to measure all four frequencies in each ear at each assessment, though 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 second analytic purpose was to evaluate whether Program-level scores were sensitive to change for children who had a second assessment using the same test. Sensitivity to change over time was coarsely evaluated using paired t-tests between 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 a paired t-test of percentile ranks as the test does not report standard scores or growth scale values. We corrected for multiple comparisons using Bonferroni’s correction.

4.1.3 Analyses: Procedure feasibility

Surveys (see Appendix 7) 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 either were in yes/no format or used 5-point
Likert scales to measure the strength of SLPs’ agreement with statements. Results are reported descriptively.

4.1.4 Results: Data usability

All regression analyses met assumptions of normality, independence, homoscedasticity, and linearity with the exception of the PLS-5 Expressive Communication models, which were significantly heteroskedastic. 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 4-5). 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. In both cases, the presence of additional factors was the only significant predictor.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Auditory Comprehension</th>
<th>Expressive Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>0.28***</td>
<td>0.235***</td>
</tr>
<tr>
<td>Better Ear Pure Tone Average (dB HL)</td>
<td>0.263</td>
<td>-0.1</td>
</tr>
<tr>
<td>Presence of additional factors affecting outcome</td>
<td>-24.13***</td>
<td>-20.79***</td>
</tr>
</tbody>
</table>

Note: * = \(p<0.05\), ** = \(p<0.01\), *** = \(p< 0.001\)

Note: BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss

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: \(F(2,24) = 0.77, p>0.05\).]
$p>0.05$: words understood: $F(2,23) = 2.45, p>0.05$; see Table 4-6]. 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 4-6: Relation of MBCDI-2 percentile ranks with predictors

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Phrases Understood</th>
<th>Words Produced</th>
<th>Words Understood</th>
<th>Gestures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>$R^2$(adj)</td>
<td>b</td>
<td>$R^2$(adj)</td>
<td>b</td>
</tr>
<tr>
<td>Model</td>
<td>0.11</td>
<td>-0.02</td>
<td>0.10</td>
<td>-0.09</td>
</tr>
<tr>
<td>Better Ear Pure Tone Average (dB HL)</td>
<td>0.07</td>
<td>-</td>
<td>0.10</td>
<td>-0.09</td>
</tr>
<tr>
<td>Presence of additional factors affecting outcome</td>
<td>-17.97*</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Presence of additional factors affecting outcome</td>
<td>12.1</td>
<td>22.44*</td>
<td>27.27**</td>
<td></td>
</tr>
</tbody>
</table>

*Note: BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss*

With regard to change over time, PLS-5 standard scores did not differ significantly between first and second assessments for either scale [Auditory Comprehension: $t(26) = 1.5623, p>0.0125$; Expressive Communication: $t(26) = -0.15823, p>0.0125$] but growth scale values increased significantly 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. Descriptively, possible declines between assessments may exist for the Words Understood and Words Produced scores but there was likely no change between assessments for Phrases Understood or Gestures (see Table 4-4).
4.1.5 Results: 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 8-12. 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 Program-level procedures either in isolation, or in conjunction with Tier 2 individual vulnerability testing procedures, was appropriate for clinical practice.

4.2 Pilot Study 2 – Tier 2 Individual Vulnerability Testing

4.2.1 Method

4.2.1.1 Procedure

The decision to participate in 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 (see Appendix 13) to identify barriers and facilitators to implementation.

4.2.1.2 Participants: Children assessed in the pilot

At the start of the pilot, SLPs from regions that our research team believed were involved in pilot study 2 flagged many children \( n=72 \) of 238 as not being involved in
the pilot (i.e., they were only including these children in the study 1 pilot). Over the course of the pilot, we became aware of a communication breakdown, after which expectations were re-communicated. As the pilot progressed, there was a trend whereby SLPs who originally indicated they were not involved in the Tier 2 pilot began to enter individual vulnerability data, however, a significant amount of missing data (57% of children in piloting regions) was observed \((n=126\) 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 GFTA-3 and CELF-P2 (see Table 4-7). We included data for all children included in the REDCap database, 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.

4.2.2 Analysis: Data usability

At the start of the pilot, SLPs from regions that our research team believed were involved in pilot study 2 flagged many children \((n=72\) of 238) as not being involved in the pilot (i.e., they were only including these children in the study 1 pilot). Over the course of the pilot, we became aware of a communication breakdown, after which
Table 4-7: Demographics of children included in pilot study 2

<table>
<thead>
<tr>
<th>Variable</th>
<th>GFTA-3 Sounds-in-Words Assessments (n=48)</th>
<th>CELF-P2 Word Structure Assessments (n=46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presence of Additional Factors</td>
<td>16</td>
<td>10</td>
</tr>
<tr>
<td>Bilateral Hearing Loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEPTA (dB HL)</td>
<td>51.47 (18.75 – 98.33)*</td>
<td>47.43 (17.25-92.5)*</td>
</tr>
<tr>
<td>Unilateral Hearing Loss</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>PTA (dB HL)</td>
<td>47.74 (26.25 – 81.25)</td>
<td>47.96 (28.75 – 83.75)</td>
</tr>
<tr>
<td>Conductive Hearing Loss</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Age at PLS Assessment (months)</td>
<td>43.54 (31 – 71)</td>
<td>48.63 (34 – 71)</td>
</tr>
<tr>
<td>Expressive Communication (Standard Score)</td>
<td>103.85 (64 – 150)</td>
<td>19.12</td>
</tr>
<tr>
<td>Auditory Comprehension (Standard Score)</td>
<td>103.34 (73 – 150)</td>
<td>16.67</td>
</tr>
<tr>
<td>Age at IVT Assessment (months)</td>
<td>44 (31 – 71)</td>
<td>48.87 (37 – 72)</td>
</tr>
<tr>
<td>GFTA-3 Sounds-in-Words (Standard Score)</td>
<td>89.32 (42 – 123)</td>
<td>42.22 (0.1 – 99)</td>
</tr>
<tr>
<td>CELF-P2 Word Structure (Percentile Ranks)</td>
<td></td>
<td>35.74</td>
</tr>
</tbody>
</table>

Note: BEPTA = better-ear pure-tone average; dB HL = decibels Hearing Loss; PTA = pure-tone average
Note: 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
Note: Standard scores are not available for the CELF-P2 subtests

expectations were re-communicated. As the pilot progressed, there was a trend whereby SLPs who originally indicated they were not involved in the Tier 2 pilot began to enter individual vulnerability data, however, a significant amount of missing data (57% of children in piloting regions) was observed (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 GFTA-3 and CELF-P2 (see Table 4-7). We included data for all children included in the REDCap database, 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.

4.2.3 Analysis: Procedure feasibility

Survey data were analyzed descriptively as in pilot study 1.

4.2.4 Results: Data usability

The proportions of children considered within, borderline, or below age expectations for each test are reported in Tables 4-8 to 4-11. 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 ($W_t(46)=0.71, p<0.05$) and CELF-P2 ($W_t(43)=0.73, p<0.05$). 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 other assessment and a lack of sensitivity/specificity data to define within/borderline/below age expectations for the MBCDI Words and Sentences form.

Table 4-8: Agreement between PLS-5 Auditory Comprehension and GFTA-3 Sounds-in-Words Subtest

<table>
<thead>
<tr>
<th></th>
<th>GFTA-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within</td>
</tr>
<tr>
<td>PLS-5</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>32</td>
</tr>
<tr>
<td>Border</td>
<td>4</td>
</tr>
<tr>
<td>Below</td>
<td>0</td>
</tr>
</tbody>
</table>
4.2.5 Results: Procedure feasibility

At the end of the Program-level pilot, 36 SLPs completed online surveys to provide feedback on the new Tier 2 procedures. 13 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 14-17.

Table 4-9: Agreement between PLS-5 Expressive Communication and GFTA-3 Sounds-in-Words Subtest

<table>
<thead>
<tr>
<th>PLS-5</th>
<th>Within</th>
<th>Border</th>
<th>Below</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within</td>
<td>31</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Border</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Below</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4-10: Agreement between PLS-5 Auditory Comprehension and CELF-P2 Words Structure Subtest

<table>
<thead>
<tr>
<th>PLS-5</th>
<th>Within</th>
<th>Border</th>
<th>Below</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within</td>
<td>17</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Border</td>
<td>0</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Below</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4-11: Agreement between PLS-5 Expressive Communication and CELF-P2 Words Structure Subtest

<table>
<thead>
<tr>
<th>PLS-5</th>
<th>Within</th>
<th>Border</th>
<th>Below</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within</td>
<td>16</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>Border</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Below</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>
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, whereas the remaining SLPs were divided between agreeing and disagreeing.

### 4.3 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 test 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 tool. 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 (Joint Committee, 2007, 2013, 2019; Moeller et al., 2013), there is limited guidance with how to accomplish the diverse assessment purposes proposed under these recommendations.

4.3.1 Data Usability

For analytic purposes, data from pilot study 1 suggest that the PLS-5 might be preferable to use for children of all ages, rather than including the MBCDI-2 at younger ages. This is because PLS-5 growth scale values were sensitive to change over time (Daub et al., 2017) and standard scores were predicted by additional factors. In this regard, the PLS-5 conformed to our predictions whereas the MBCDI-2 did not, 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 (<12 months, as a group). 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., 2007; Feldman et al., 2007). 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.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) are higher than what is typically reported in outcome studies (e.g., Tomblin et al., 2015) and approximate a normal distribution, which has a standard score mean of 100
and standard deviation of 15. In our data, children without additional factors scored (as a group) at a mean of 100.92 \((SD=20.5)\) on the Expressive Communication scale and at a mean of 98.98 \((SD=19.81)\) on the Auditory Comprehension scale (see Appendix 18). 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 Appendices 2-6). 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 are preliminary evidence that the IHP is achieving their goal of ameliorating the impact of inconsistent auditory access on spoken language outcomes. However, we remain cautious in this interpretation. 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 province-wide. Additionally, 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 (Cupplies et al., 2014). It is also possible that some performance factors (e.g., children’s 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.
4.3.2 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 is insufficient to make recommendations for EHDI programs and to determine whether tests provide unique predictive information, SLPs’ feedback indicates that valuable 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 relations between the proposed measures in the individual vulnerability testing procedure.

4.3.3 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 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., accepted; Messick, 1998). 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 suggested 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 with SLPs the tests we selected for Tier 2 monitoring based on our previous analysis (Daub & Oram Cardy, 2021), and the rationale for monitoring key areas of vulnerability in children who are DHH. This leaves SLPs free to use the recommended Tier 2 tools when they identify a need in clinical practice, rather than mandating it program-wide at this point.

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.

4.4 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


Development inventories at ages one and two years. *Child Development, 71*(2), 310–322. doi.org/10.1111/1467-8624.00146


Chapter 5

5 What do Speech-Language Pathologists want to know when assessing early vocal development in children who are (D)eaf/hard-of-hearing?

Early linguistic experiences influence infants’ processing of future linguistic experiences and lays the foundation for later language outcomes (e.g., Jansson-Verkasalo et al., 2010; Johnson & Jusczyk, 2001; Kuhl et al., 2008; Moon, Lagercrantz & Kuhl, 2012; Nazzi & Ramus, 2003; Tsao et al., 2004; Thiessen & Saffran, 2007; Werker & Tees, 1984). Permanent childhood hearing loss reduces infants’ and children’s experience with spoken language (Moeller & Tomblin 2015) and children who are deaf/hard-of-hearing (CDHH) are at increased risk for poorer overall spoken language outcomes than their typically-hearing peers (Joint Committee of Infant Hearing, 2013; Moeller, 2000; Nelson et al., 2008; Patel & Feldman, 2011). Hearing loss itself is not a language learning disorder, but a sensory disorder that impoverishes the child’s linguistic environment with cascading effects on language learning and development. When hearing loss’ impact on CDHH’s language environment is adequately mitigated, it is expected that CDHH can acquire language, either signed or spoken, within the expectations established for their same-aged peers (Joint Committee of Infant Hearing, 2013).

Early Hearing Detection and Intervention (EHDI) programs are committed to mitigating the impact of hearing loss on early linguistic experiences through the early identification of hearing loss and timely, comprehensive, supports to families and children in order to create rich (signed or spoken) language learning environments. Within EHDI programs, the choice to pursue signed or spoken language is the family’s
(Moeller et al., 2013), with support from EHDI service providers (e.g., sign-language consultants, audiologists, speech-language pathologists).

Monitoring language development has been argued to be crucial for identifying CDHH who are showing signs of difficulty in language learning so that intervention efforts, either technological or behavioural, can be tailored (Joint Committee of Infant Hearing, 2013; Moeller et al., 2013). Beyond intervention planning and family counselling, language outcome monitoring is also recommended to inform broader curricular and resource decisions at the level of the overall EHDI program. Of interest in the present paper is the measurement of early vocal development for families who choose to teach their child a spoken language. For the purposes of the present paper, vocal development is defined as including the early vocalizations associated with protophone development (Oller, 2000), including canonical babble, as well as a child’s repertoire of speech sounds, syllable shapes, and syllable complexity (Moeller et al., 2007). Differences in vocal development, particularly canonical babble, have been routinely documented between CDHH and children with typical hearing (Ambrose et al., 2016; Iyer & Oller, 2008; Moeller et al., 2007; Oller, 2000). Prolonged delays in canonical babble, and reductions in syllable complexity, have been demonstrated to be predictive of ongoing language delays later in development (Moeller et al., 2007). Vocal development assessments, therefore, have the potential to inform intervention planning and goal setting.

There are, however, very few vocal development tests available. None of the existing tools report normative scores to determine whether a child’s vocal development is within age-expectations, and all are missing some elements of validity evidence that
would be necessary to recommend them for implementation. Further, none have sufficient validity evidence to recommend their implementation specifically in an EHDI context.

5.1 Existing vocal development tests

5.1.1 LittlEARs Early Speech Production Questionnaire® (LEESPQ®)

The LEESPQ is a 27-item, yes/no, parent-completed questionnaire intended to be used with children between birth and 18 months of age. The LEESPQ was originally developed in German (Koşaner et al., 2014), and evidence of the LEESPQ’s ability to accurately capture the early spoken language development of children who are typically hearing and developing has been appraised in several languages (Keilmann et al., 2018; Wachtlin et al., 2017), including English (Daub, Oram Cardy, et al., 2019). Results from these studies have demonstrated that LEESPQ scores are significantly related to age and scores on a parent report measure of early spoken language (the Receptive-Expressive Emergent Language Test – Third Edition; Bzoch, League & Brown, 2003; Daub, Oram Cardy, et al., 2019), and are unrelated to sex or multilingual status (Keilmann et al., 2018). Unlike other vocal development assessments, work with the LEESPQ has not yet explored whether it is related to either clinical, or acoustic, analysis of vocal behaviours. However, the LEESPQ is limited in capturing early vocal behaviours prior to the onset of canonical babbling, such as marginal babble and phonation, due to a lack of items that specifically measure these abilities. The items on the LEESPQ that aim to measure behaviours earlier than canonical babble did not contribute significant amounts of information to the child’s total score (Daub, Oram Cardy, et al., 2019). Performance characteristics have not been established for the LEESPQ with CDHH and for the
LEESPQ’s sensitivity to differences in trajectory, or ability to predict later language outcomes.

5.1.2 Vocal development Landmarks Interview (VDLI)

The VDLI (Ambrose et al., 2016; Moeller et al., 2019) is unique relative to other parent-report measures of vocal development in that it uses digital audio recordings of natural infant vocalizations to support parent responding. In the interview, parents are asked to listen to audio examples or developmental contrasts of vocal behaviors and report on whether their infant makes sounds similar to the model. The VDLI includes developmental behaviors expected in the age range of 6 to 21 months and takes between 20 – 30 minutes to administer (Ambrose et al., 2016; Moeller et al., 2019). An experimental version of the VDLI was used in the Outcomes of Childhood Hearing Loss study (Tomblin et al., 2015), and vocal development stage was significantly related to children’s age and hearing status, with CDHH scoring lower than children with normal hearing (Ambrose et al., 2016). In 2019, Moeller et al. reported validity evidence for a revised version of the VDLI in a sample of typically developing children between 6 and 21 months of age. These results indicated that the VDLI is significantly correlated with scores on the Communication and Symbolic Behaviour Scales - Developmental Profiles (Wetherby & Prizant, 2002) and items on individual subscales were found to be internally consistent. Future work is planned to look at the levels of agreement between parents and researchers completing the VDLI and the current data are limited by having been collected on a sample of children born to families with high socio-economic status relative to the broader U.S. population (Moeller et al., 2019).
5.1.3 Infant Monitor of Vocal Productions (IMP)

The IMP was not originally designed as a clinical assessment tool. Instead, it was developed as an educational resource to teach parents of babies (younger than 12 months) about stages of vocal development and attune them to their baby’s current vocal abilities, with the goal that parents and service providers could make informed decisions about their child’s audiological management (Cantle Moore, 2014). Normative data has been collected using the IMP for infants with normal hearing between 3 and 14 months of age, including normative curves and percentiles (Cantle Moore & Colyvas, 2018). The IMP has been demonstrated to be unrelated to maternal education, gender, or multilingual status (Cantle Moore & Colyvas, 2018). Additionally, preliminary pilot data with 9 children who were binaurally fitted with hearing aids and 9 children with normal hearing demonstrated that children who wore hearing aids scored lower on the IMP than children with normal hearing, although their rate of growth in scores over time was comparable (Cantle Moore, 2014). These results have not been confirmed with larger samples.

5.1.4 Prelexical Infant Scale Evaluation (PRISE)

The PRISE relies on parent interview elicitation to collect information about children’s pre-lexical vocal behavior in the child’s everyday context, using 11 probes (Kishon-Rabin et al., 2005). The 11 probes were developed to capture vocal development milestones in children with normal hearing (Kishon-Rabin et al., 2005). The PRISE has been used as a predictor of functional hearing (Kishon-Rabin et al., 2009) and auditory skills development following cochlear implantation in children with hearing loss between 8 and 23 months of age (Kishon-Rabin et al., 2005). At present, there have been no data
published establishing the PRISE’s relation with other measures of early spoken language or vocal development.

5.2 What Validity Evidence is Needed?

All vocal development tests are currently limited with respect to the validity evidence that is available to support their use. There is similarly a lack of evidence for the clinical feasibility of each of the assessments in an EHDI context. EHDI systems are complex to implement and are inherently interprofessional – they require intensive intervention services at various levels, and coordination across multiple service providers who are required to collect clinical information that informs not only their clinical practice, but the practice of other professionals. For instance, information about a CDHH’s speech sound production abilities is useful to speech-language pathologists (SLPs) for selecting intervention targets, but speech sound production abilities can also point to important information about the child’s auditory skills that can be modified by audiologists (e.g., non-linear frequency lowering).

All currently available vocal development tools are in phases of development where additional validity evidence is needed prior to their uptake in clinical practice. However, all assessment tools, commercially available or experimental, will necessarily lack some validity evidence in that not all tools are equipped for to answer all assessment questions, and the properties that make a test appropriate for some questions makes them less appropriate for others (Peña, Spaulding & Plante, 2006). Moreover, the issue of what validity evidence is needed, what is desirable, and what is superfluous needs consideration. As a field, speech-language pathology has grappled with shifting demands in psychometric best practice and test development for decades. In the 1980s, reviews of
standardized assessments identified that there were very few child language assessments containing any information regarding diagnostic accuracy (McCauley & Swisher, 1984a). Advocacy work begot changes in tests, and improvements in diagnostic accuracy reporting began in the 1990s (Plante & Vance, 1994) and continued through the 2000s (Flipsen & Ogiela, 2015; Denman et al., 2017). However, the bar has once again shifted and SLPs report requiring more information from their norm-referenced tests than current tests are validated to support (Kerr et al., 2003) and are using norm-referenced tests outside the purposes for which the tests were initially validated.

Modern conceptualizations of validity adopt an argument-based approach to validity (American Educational Research Association, 2014; Messick, 1993). Under this perspective, there are not distinct categories of validity (i.e., face, content, construct). Instead, there are only different types of evidence that fall under the broader construct of validity. In this view, the decisions SLPs make are the object of validation, not the tests themselves. Therefore, the extent to which a test is or is not appropriate is defined by relating statistical evidence to individual SLPs’ decisions. For instance, including children with a disorder in the composition of normative groups improves a test’s ability to determine a child’s ability relative to other children with speech and language disorders, but lowers the test’s ability to detect whether or not a child is below age-expectations compared to same-aged, typically developing peers (Peña, Spaulding & Plante, 2006). These connections between decisions and appropriate evidence are not necessarily always explicit in examiner manuals. Daub, Skarakis-Doyle, et al. (2019) highlighted the disconnect between modern conceptualizations of validity, how commercially available norm-referenced tests are developed, and how SLPs used
assessments. They hypothesized that including SLPs’ perspectives early into the test development process and relating validity evidence to their decision making could improve commercially available tests as well as evidence-based assessment practice.

One aim of our research program is to support Canadian SLPs working in a publicly funded EHDI program to identify the best test for tracking the early vocal development of CDHH. In the present paper, we adopted the position proposed by Daub, Skarakis-Doyle et al. (2019) that a necessary first step in this process was to identify the clinical decisions about vocal development that SLPs identified as most important to their clinical practice. This will enable the future step of mapping the decisions SLPs need to make onto validity evidence of existing vocal development tests (or to develop new evidence or tests) so that those tests most appropriate for making the desired clinical decisions can determined. Specifically, in partnership with Ontario’s Infant Hearing Program (described below), we were planning future studies to collect additional validity evidence for the LEESPQ and the VDLI to identify whether either of the two tests was equipped to support the SLPs’ assessment purposes. We expected that by surveying SLPs working in this EHDI program, we would be able to: a) inform efforts to conduct new validity investigations of existing vocal development tests, b) inform design of new vocal development tests intended for EHDI contexts, and c) demonstrate an approach to test design and validation that incorporates SLPs’ perspectives.

5.3 Study Purpose

The present study was part of a larger series of quality improvement projects conducted with the Ontario Infant Hearing Program (IHP). The IHP is a publicly funded EHDI program that provides universal newborn hearing screening to all children in
Ontario, and family-centered supports to all children identified with permanent childhood hearing loss from identification to their transition to school-based services (typically by the age of 6 years in Ontario). Intervention supports are determined by the family, and care-plans can include sign-language supports, speech-language pathology services, auditory verbal therapy, and audiological intervention. Within the Ontario context, CDHH who receive cochlear implants are managed by a separate program. Therefore, SLPs providing services to children in the IHP are typically providing services to CDHH with some degree of residual hearing, whose losses may be mild to profound, and who are (typically) amplified with hearing aids. At the time of this study, the IHP was in the process of developing and implementing a spoken language outcome monitoring protocol, and our group was involved in consulting to the IHP during this process. Because there is a lack of compelling evidence to guide the selection of one vocal development test over another within the program, the IHP wanted to conduct a series of projects to support the selection of a vocal development assessment tool for children younger than 22 months of age. We selected this age range to support our future planned studies evaluating the LEESPQ and the VDLI. At the time of our study design, validity evidence was reported for the VDLI for children up to 21 months of age.

The present study was our first step in supporting the IHP in selecting a vocal assessment tool. We initiated the study to understand SLPs’ vocal development assessment purposes, that is, the clinical decisions they seek to make based on their assessment of early vocal development, as well as barriers to assessment of vocal development that might exist in clinical practice and influence the selection of one tool over another. The primary purpose of this study was to identify the assessment purposes
that are the highest priorities to SLPs that could then serve as a basis for designing future
validity studies. We expect that understanding SLPs’ assessment purposes in this way
will enable us to conduct clinically relevant validity projects to support the eventual
implementation of new tools into clinical practice as predicted by Daub, Skarakis-Doyle,
et al. (2019). Our secondary purpose was to understand the barriers to vocal development
assessment of SLPs practicing in the Ontario IHP. Information about the barriers that
SLPs experience in assessing vocal development were expected to inform future tool
design projects by identifying potential modifications to the tool (e.g., reducing test
length of time to assess is considered a major barrier) that would support the clinical
uptake of the tool.

5.4 Methods
5.4.1 Ethical Approval

Data collection for this study was completed as part of a larger government
Program Evaluation and Quality Improvement project with the Ontario Ministry of
Children, Community and Social Services that was reviewed by the Western University
Research Ethics Board (REB). The REB considered the project 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 it was not considered to fall under the purview of
the REB.

5.4.2 Participants

The IHP does not employ its own team of SLPs and before school entry, CDHH in
Ontario who are learning spoken language access speech-language pathology services
through the Ontario Preschool Speech and Language Program. In the Preschool Speech
and Language Program, over 400 SLPs are employed in 29 regions across the province.
Whereas the IHP provides services to approximately 11,000 CDHH, SLPs in the
Preschool Speech and Language program provide services to more than 60,000 children
between birth and school entry with speech, language, or communication needs (e.g.,
children with developmental language disorder, late talkers, autism, etc.), not just CDHH.
Across regions, there are differences in how SLP services are allocated to children
enrolled in the Ontario IHP. In some regions, certain SLPs are designated to support all
CDHH in that region, whereas in others, any SLP may see a child with permanent
hearing loss (along with children with a variety of other needs). Due to the complexities
and regional variability in resource allocation, the exact number of SLPs providing
services to IHP children across Ontario is unknown. Additionally, how SLP services are
allocated may change over time in response to staffing and caseload needs. Therefore,
although there are 400 SLPs employed by the program, it is not the case that all 400 SLPs
provide services a) to children with hearing loss, or b) provide services to children
younger than 22 months. Because of the variability in how SLP caseloads are managed
across regions, and time, the percentage of SLPs who would be eligible to participate in
our survey is unknown. Regional managers were asked to forward an invitation to
participate in the survey to SLPs in their region who provide services to children from the
IHP, but we do not know the exact number of SLPs who were invited.

One hundred and two SLPs who provide services to CDHH responded to the online
anonymous survey. Of these SLPs, 74 reported having children with permanent hearing
loss younger than 22 months on their caseload and deemed eligible to include for
analysis. Fifty-nine (79.73%) of the eligible surveys contained complete responses. In one instance, a survey respondent indicated that they did not believe the survey adequately captured their experiences, so data for this respondent were excluded in our analyses. We report the data for a final 58 respondents.

As a group, our participants were highly experienced SLPs and many had advanced training in supporting CDHH. SLPs included in our final analyses had a mean of 16.4 years (SD=7.57, range: 3-34 years) of experience working as a SLP and 15.04 (SD=7.04, range: 2.5-34) years providing services to children enrolled in the Preschool Speech and Language Program. Years of experience, both as a SLP and as a clinician working within the Preschool Speech and Language Program, did not significantly violate Shapiro-Wilk’s normality test (years as a SLP: W=0.98, p >0.05; years working in the program: W=0.97, p>0.05) suggesting that experience, broadly conceptualized, was normally distributed. 34 (59%) SLPs reported that they provide auditory verbal services. Within the program, auditory verbal services may be provided by a certified Auditory Verbal Therapist or by SLPs who have completed additional professional development at a designated IHP training site but are not certified as Auditory Verbal Therapists.

5.4.3 Online Survey

Survey design was informed by The Revised Ottawa Model of Research Use (OMRU; Graham & Logan, 2004). The OMRU is a prescriptive model of implementation science, where implementation interventions are advised to Assess, Monitor, and Evaluate aspects of an evidence-based innovation, potential adopters and the practice context (see Appendix 19). The survey used in this study (see Appendix 20) was conceptualized to Assess aspects of the evidence-based innovation (i.e., clinical decisions
that the vocal development tools should be validated to support) as well as aspects of potential adopters (SLPs) and the practice context (publicly funded Infant Hearing and Preschool Speech and Language programs).

The first set of questions in the survey was designed to understand barriers to vocal assessment from the perspective of SLPs (potential adopters) and their practice context using the components of the OMRU as a framework. This section also contained questions pertaining to barriers to assessing first words and early lexical development in young children. These questions were included because first words are another domain of spoken language that is particularly vulnerable in children with permanent hearing loss (Moeller et al., 2007) and for which there are more commercially available norm-referenced tests (e.g., *MacArthur-Bates Communicative Development Inventories*; Fenson et al., 2007). Asking about first word assessment was expected to highlight barriers to vocal development that may be more pronounced than simply the complexity related to assessing young children. However, barriers related to first words were not our primary focus for analysis. Questions in this section were modelled on surveys originally designed by Moodie and colleagues (2011) to understand pediatric audiologists’ perceptions of a new auditory outcome monitoring procedure and then adapted through the lens of the OMRU in order to understand barriers to implementing a spoken language outcome monitoring procedure in a publicly funded EHDI program (Cunningham et al., 2019).

A second section was dedicated to understanding the assessment decisions that SLPs believe are important to their clinical practice as well as their current assessment practices and barriers to vocal development assessment. We collected data in two ways:
first, we attempted to collect open-ended statements to support a planned secondary concept mapping analysis, and second, we collected quantitative data where SLPs rated the importance of various purposes. The statement generation section of the survey contained a series of prompts (e.g., “In my clinical practice, I use the results of a child with permanent hearing loss’ vocal development (re)assessment to___”) intended to elicit single, full sentence, statements. The purpose of these prompts was to collect data for structured conceptualization and concept mapping analysis. Structured conceptualization is a mixed methods technique designed to capture perspectives from groups of stakeholders and concept mapping refers to the visualization (“mapping”) of the results from structured conceptualization to support decision making (Trochim & Kane, 2005).

The process involves brainstorming activities, which can be conducted individually or in groups, in person, or remotely followed by having participants sort statements generated by their peers into like categories, to identify relationships between statements (Kane & Trochim, 2007). Structured conceptualization has been used in a variety of disciplines (Trochim & Kane, 2005), most recently in speech-language pathology to identify solutions to assessment barriers in a publicly funded intervention program (Kwok et al., 2020). However, sorting and interpreting concept mapping statements requires that the statements contain only one idea and are presented in full sentence form (Kane & Trochim, 2007) – criteria that the statements SLPs provided in their survey responses did not fulfill. For these reasons, concept mapping of the statements provided by these responses was deemed to be inappropriate, and the present paper reports the results from the quantitative questions asked in the next section. Because we were relying on a remote brainstorming process, we were aware that there was a possibility that the responses SLPs
would provide might not conform to the criteria for concept mapping. To address this concern, we incorporated a second set of questions about assessment purposes quantitatively.

In our quantitative questions, we asked SLPs to rate the importance of 15 assessment purposes on a 5-point Likert scale from “not at all important” to “very important”. The 15 assessment purposes were developed by the first and last authors who have clinical experience in speech-language pathology (O.D. as a student-clinician, and J.O.C. as a registered SLP). The assessment purposes were intentionally designed to capture a range of purposes, such as diagnosis, goal setting, and progress monitoring.

During survey design, we speculated that SLPs might reasonably report that all 15 assessment purposes are important to their clinical practice, which would not support our goal of prioritizing assessment purposes for future exploration. Therefore, SLPs were also asked to identify 5 assessment purposes that would be the most important for a vocal development assessment to be equipped to answer. Finally, because the 15 statements were generated by the authors and not the clinicians themselves, we included a final two questions asking respondents to indicate if there were other clinically important assessment questions they have that were not included in our list, and to specify any additional questions they have that were not included.

5.5 Analysis
5.5.1 Vocal Development Assessment Purposes

SLPs’ 5-point Likert scale ratings of each purpose’s importance were evaluated descriptively. Purpose rankings were evaluated with respect to identifying which purposes were flagged the most frequently as belonging in SLPs’ “Top 5” assessment
purposes. In order to identify whether an assessment purpose was endorsed by the majority of SLPs, we also examined whether any assessment purpose was identified as belonging in more than 50% of SLPs “Top 5”.

5.5.2 Barriers to Vocal Development Assessment

Barriers to vocal development were similarly evaluated descriptively. To date, there is limited published guidance on how to identify the level of agreement that indicates whether an item acts as a barrier in clinical practice. However, previous work designed to identify actionable items to target in implementation interventions in the IHP pragmatically used a criterion of less than 60% agreement with an item (Cunningham et al., 2019). Although there is no evidence to suggest that applying a 60% criterion identifies barriers which are more, or less, influential in implementation, using a 60% criterion was thought to correspond to a reasonable majority of SLPs. For our purposes of supporting the IHP, we categorized items as barriers if fewer than 60% of SLPs responded positively, and we considered how changing the criterion we applied would influence our decision making. All items were positively worded and reverse keying was not required.

5.6 Results

5.6.1 Assessment Purposes

We considered our 15 pre-developed statements to be representative of clinicians’ perspectives if 90% or greater responded ‘No’ to our question asking if they had any additional purposes not included in the survey. This criterion was established based on percent agreement criteria used in Delphi studies (which typically range between 50-80%
agreement to be defined as consensus) and is in line with stricter criteria that have been adopted by researchers working in policy making decisions (i.e., Cunningham, et al., 2019b). 52/58 (90%) of clinicians indicated they did not have additional assessment purposes. Participants who indicated they did have additional purposes were asked to list them. One SLP stated they didn’t have additional assessment questions but provided additional practice context, and one SLP listed a broader question about the availability of vocal development assessments, rather than a clinical assessment purpose. These two ‘No’ responses were judged by the research team to not represent the SLPs’ opinions about our 15 generated statements, and 54/56 (96%) of clinicians had no further questions to add, meeting our criterion of 90%.

Each of the 15 assessment purposes were rated as “Somewhat” or “Very” important by the large majority of SLPs (> 90%; see Table 5-1) indicating that SLPs approach vocal development assessments with numerous purposes. There was less clarity in which assessment purposes were identified as the most important. All purposes were rated as belonging in some SLPs’ “Top 5” assessment purposes (see Table 5-2). However, only three assessment purposes were prioritized by more than 50% of SLPs: “Does the child’s level of vocal development indicate that the child is having more problems with speech development than expected based on their hearing loss?”; “Does the child’s level of vocal development indicate the child is having more problems with language learning than expected based on their hearing loss?”; “Has the child acquired new vocal development abilities since their last visit?” Two of these purposes primarily correspond to using tests for differential diagnosis and the third primarily relates to measuring progress.
Table 5-1: SLPs beliefs about assessment purposes' importance

<table>
<thead>
<tr>
<th>Question</th>
<th>Very Unimportant</th>
<th>Somewhat Unimportant</th>
<th>Neutral</th>
<th>Somewhat Important</th>
<th>Very Important</th>
<th>Mode (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the child’s vocal development within age-expectations compared to children their age who are typically developing and have typical hearing?</td>
<td>0 (0%)</td>
<td>3 (5%)</td>
<td>2 (35%)</td>
<td>24 (41%)</td>
<td>29 (50%)</td>
<td>5 (2-5)</td>
</tr>
<tr>
<td>Is the child’s vocal development within expectations for children with similar levels of hearing loss?</td>
<td>1 (2%)</td>
<td>0 (0%)</td>
<td>6 (10%)</td>
<td>18 (31%)</td>
<td>33 (57%)</td>
<td>5 (1-5)</td>
</tr>
<tr>
<td>Is the child’s vocal development within expectations for children with similar amplified hearing levels?</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>3 (5%)</td>
<td>18 (31%)</td>
<td>37 (64%)</td>
<td>5 (3-5)</td>
</tr>
<tr>
<td>Has the child’s vocal development improved, relative to their same-aged peers, since their last visit?</td>
<td>0 (0%)</td>
<td>1 (2%)</td>
<td>5 (9%)</td>
<td>12 (21%)</td>
<td>40 (69%)</td>
<td>5 (2-5)</td>
</tr>
<tr>
<td>Has the child acquired new vocal development abilities since their last visit?</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (2%)</td>
<td>6 (10%)</td>
<td>51 (88%)</td>
<td>5 (3-5)</td>
</tr>
<tr>
<td>Has the child’s vocal development fallen behind their same-aged peers since their last visit?</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>3 (5%)</td>
<td>13 (22%)</td>
<td>42 (72%)</td>
<td>5 (3-5)</td>
</tr>
<tr>
<td>Has the child’s vocal development plateaued or not changed since their last visit?</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>1 (2%)</td>
<td>7 (12%)</td>
<td>50 (87%)</td>
<td>5 (3-5)</td>
</tr>
<tr>
<td>Does the child’s level of vocal development indicate the child is having more problems with language learning than expected based on their hearing loss?</td>
<td>0 (0%)</td>
<td>1 (2%)</td>
<td>5 (9%)</td>
<td>15 (26%)</td>
<td>37 (64%)</td>
<td>5 (2-5)</td>
</tr>
<tr>
<td>Does the child’s level of vocal development indicate that the child is having more problems with speech development than expected based on their hearing loss?</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>2 (3%)</td>
<td>11 (19%)</td>
<td>45 (78%)</td>
<td>5 (3-5)</td>
</tr>
<tr>
<td>Does the child’s level of vocal development indicate that the child needs more speech and language therapy than they are currently receiving?</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>10 (17%)</td>
<td>16 (28%)</td>
<td>32 (55%)</td>
<td>5 (3-5)</td>
</tr>
<tr>
<td>What stage of vocal development has the child mastered?</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>3 (5%)</td>
<td>13 (22%)</td>
<td>42 (72%)</td>
<td>5 (3-5)</td>
</tr>
<tr>
<td>What stage of vocal development is emerging?</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>3 (5%)</td>
<td>15 (26%)</td>
<td>40 (69%)</td>
<td>5 (3-5)</td>
</tr>
</tbody>
</table>
What speech sounds would be appropriate goals for the child?  
0 (0%)  1 (2%)  2 (3%)  13 (22%)  42 (72%)  5 (2-5)

Which syllable shapes would be appropriate goals for the child?  
0 (0%)  0 (0%)  1 (2%)  12 (21%)  45 (78%)  5 (3-5)

Which words would be appropriate goals for the child?  
0 (0%)  0 (0%)  4 (7%)  16 (28%)  38 (66%)  5 (3-5)

Table 5-2: SLPs prioritization of assessment purposes

<table>
<thead>
<tr>
<th>Assessment Purposes</th>
<th>In my top 5 (%)</th>
<th>Not in my top 5 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Is the child’s vocal development within age-expectations compared to children their age who are typically developing and have typical hearing?</strong></td>
<td>26 (45%)</td>
<td>32 (55%)</td>
</tr>
<tr>
<td>Is the child’s vocal development within expectations for children with similar levels of hearing loss?</td>
<td>13 (22%)</td>
<td>45 (78%)</td>
</tr>
<tr>
<td>Is the child’s vocal development within expectations for children with similar amplified hearing levels?</td>
<td>19 (33%)</td>
<td>39 (67%)</td>
</tr>
<tr>
<td>Has the child’s vocal development improved, relative to their same-aged peers, since their last visit?</td>
<td>12 (21%)</td>
<td>45 (79%)</td>
</tr>
<tr>
<td><strong>Has the child acquired new vocal development abilities since their last visit?</strong></td>
<td>37 (64%)</td>
<td>21 (36%)</td>
</tr>
<tr>
<td>Has the child’s vocal development fallen behind their same-aged peers since their last visit?</td>
<td>10 (17%)</td>
<td>48 (83%)</td>
</tr>
<tr>
<td>Has the child’s vocal development plateaued or not changed since their last visit?</td>
<td>13 (23%)</td>
<td>45 (77%)</td>
</tr>
<tr>
<td><strong>Does the child’s level of vocal development indicate the child is having more problems with language learning than expected based on their hearing loss?</strong></td>
<td>29 (50%)</td>
<td>29 (50%)</td>
</tr>
<tr>
<td><strong>Does the child’s level of vocal development indicate that the child is having more problems with speech development than expected based on their hearing loss?</strong></td>
<td>37 (64%)</td>
<td>21 (36%)</td>
</tr>
<tr>
<td>Does the child’s level of vocal development indicate that the child needs more speech and language therapy than they are currently receiving?</td>
<td>10 (17%)</td>
<td>48 (82%)</td>
</tr>
<tr>
<td>What stage of vocal development has the child mastered?</td>
<td>17 (30%)</td>
<td>41 (70%)</td>
</tr>
<tr>
<td>What stage of vocal development is emerging?</td>
<td>19 (33%)</td>
<td>39 (67%)</td>
</tr>
<tr>
<td><strong>What speech sounds would be appropriate goals for the child?</strong></td>
<td>21 (36%)</td>
<td>37 (64%)</td>
</tr>
<tr>
<td>Which syllable shapes would be appropriate goals for the child?</td>
<td>11 (19%)</td>
<td>47 (81%)</td>
</tr>
<tr>
<td><strong>Which words would be appropriate goals for the child?</strong></td>
<td>18 (31%)</td>
<td>40 (69%)</td>
</tr>
</tbody>
</table>

*Note: Items in bold are the assessment purposes that SLPs most commonly reported as belonging to their “Top 5” assessment purposes.

5.6.2 Barriers to Vocal Assessment

SLPs reported no barriers to the assessment of first words using Cunningham’s (2019) conservative definition of a barrier, and in all cases more SLPs agreed or strongly
agreed with statements pertaining to first words than they did with statements pertaining to vocal development. The barriers that SLPs report, therefore, appear to be specific to assessing vocal development rather than assessing children younger than 22 months more generally. Three barriers to assessing vocal development were reported using a 60% criterion: two related to economic barriers, and a third relating to knowledge to support interpreting assessment results (see Table 5-3). In our data, applying a < 50% criterion does not significantly change the interpretation of results. The economic barriers would also meet a more liberal definition of a barrier of < 50% agreement (that is, the majority of SLPs do not agree with the statement). 80% or more of SLPs agreed with all current practice items; 100% of SLPs agreed with items relating to attitude; and 80% or more of SLPs agreed with all cultural/social items. Therefore, no cultural/social, attitudinal, or current practice barriers were reported. Despite reporting barriers to appropriate assessment tools, the majority of SLPs reported regularly assessing vocal development in children younger than 22 months on their caseload.

5.7 Discussion

This project had two objectives: first, to identify the assessment purposes that are the most important to SLPs providing services to CDHH in the IHP; and second, to identify any barriers to vocal development assessment that would influence future efforts to implement vocal development assessments in the IHP. For the IHP specifically, we were able to identify purposes that any recommended vocal development tool should fulfill (differential diagnosis, measuring, progress, and determining whether children are performing comparable to their hearing peers) to guide future validation work.
### Table 5-3: Barriers to vocal development assessment

<table>
<thead>
<tr>
<th>Potential Adopters</th>
<th>Never (%)</th>
<th>Seldom (%)</th>
<th>Sometimes (%)</th>
<th>Frequently (%)</th>
<th>Always (%)</th>
<th>Mode (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current practice</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How often do you assess the vocal development of any child (with or without permanent hearing loss) on your caseload when they are younger than 22 months?</td>
<td>0 (0%)</td>
<td>2 (3%)</td>
<td>6 (10%)</td>
<td>24 (41%)</td>
<td>26 (45%)</td>
<td>5 (2-5)</td>
</tr>
<tr>
<td>How often do you assess the first words of any child (with or without permanent hearing loss) on your caseload when they are younger than 22 months?</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>2 (3%)</td>
<td>12 (21%)</td>
<td>44 (76%)</td>
<td>5 (3-5)</td>
</tr>
<tr>
<td>How often do you assess the vocal development of children with permanent hearing loss who are younger than 22 months?</td>
<td>0 (0%)</td>
<td>6 (10%)</td>
<td>4 (7%)</td>
<td>19 (33%)</td>
<td>29 (50%)</td>
<td>5 (2-5)</td>
</tr>
<tr>
<td>How often do you assess the first words of children with permanent hearing loss who are younger than 22 months?</td>
<td>0 (0%)</td>
<td>4 (7%)</td>
<td>3 (5%)</td>
<td>7 (12%)</td>
<td>44 (76%)</td>
<td>5 (2-5)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attitudes</th>
<th>Strongly Disagree (%)</th>
<th>Disagree (%)</th>
<th>Neither Agree nor Disagree (%)</th>
<th>Agree (%)</th>
<th>Strongly Agree (%)</th>
<th>Mode (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is within my scope of practice as a SLP to assess the vocal development of children who have permanent hearing loss who are younger than 22 months</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>21 (36%)</td>
<td>37 (64%)</td>
<td>5 (4-5)</td>
</tr>
<tr>
<td>It is within my scope of practice as a SLP to assess the first words of children who have permanent hearing loss who are younger than 22 months</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>14 (24%)</td>
<td>44 (76%)</td>
<td>5 (4-5)</td>
</tr>
<tr>
<td>Assessing the vocal development of children with permanent hearing loss who are younger than 22 months provides me with important information</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>24 (41%)</td>
<td>34 (59%)</td>
<td>5 (4-5)</td>
</tr>
<tr>
<td>Assessing the vocal development of children with permanent hearing loss who are younger than 22 months provides families with important information</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>3 (5%)</td>
<td>24 (41%)</td>
<td>31 (54%)</td>
<td>5 (3-5)</td>
</tr>
<tr>
<td>Assessing the vocal development of children with permanent hearing loss who are younger than 22 months provides audiologists with important information</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>5 (9%)</td>
<td>29 (50%)</td>
<td>24 (41%)</td>
<td>5 (3-5)</td>
</tr>
</tbody>
</table>

**Knowledge/skill**
I have the knowledge I need to conduct an appropriate vocal development assessment of a child with permanent hearing loss who is younger than 22 months

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neither Agree nor Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Mode (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (2%)</td>
<td>7 (12%)</td>
<td>13 (22%)</td>
<td>29 (50%)</td>
<td>8 (14%)</td>
<td>4 (1-5)</td>
</tr>
</tbody>
</table>

I have the knowledge I need to conduct an appropriate first words assessment of a child with permanent hearing loss who is younger than 22 months

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neither Agree nor Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Mode (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (2%)</td>
<td>4 (7%)</td>
<td>8 (14%)</td>
<td>24 (41%)</td>
<td>21 (36%)</td>
<td>4 (1-5)</td>
</tr>
</tbody>
</table>

I have the knowledge I need to appropriately interpret the results of a vocal development assessment of a child with permanent hearing loss who is younger than 22 months

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neither Agree nor Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Mode (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (2%)</td>
<td>10 (17%)</td>
<td>14 (24%)</td>
<td>26 (45%)</td>
<td>7 (12%)</td>
<td>4 (1-5)</td>
</tr>
</tbody>
</table>

I have the knowledge I need to appropriately interpret the results of a first words assessment of a child with permanent hearing loss who is younger than 22 months

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neither Agree nor Disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Mode (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (0%)</td>
<td>5 (9%)</td>
<td>7 (12%)</td>
<td>30 (52%)</td>
<td>16 (27%)</td>
<td>4 (2-5)</td>
</tr>
</tbody>
</table>

### Practice Environment

#### Culture/social

I believe that other SLPs regularly assess the vocal development of children who have permanent hearing loss who are younger than 22 months

| Practice Environment |
|----------------------|---------------------|
| Strongly Disagree (%)| Disagree (%)        |
|                      | Neither Agree nor Disagree (%) | Agree (%) | Strongly Agree (%) | Mode (range) |
| 0 (0%)               | 1 (2%)              | 9 (15%) | 30 (52%) | 18 (31%) | 4 (2-5) |

I believe that other SLPs regularly assess the first words of children who have permanent hearing loss who are younger than 22 months

| Practice Environment |
|----------------------|---------------------|
| Strongly Disagree (%)| Disagree (%)        |
|                      | Neither Agree nor Disagree (%) | Agree (%) | Strongly Agree (%) | Mode (range) |
| 0 (0%)               | 0 (0%)              | 4 (7%) | 25 (43%) | 29 (50%) | 5 (3-5) |

In my opinion, families of children who have permanent hearing loss who are younger than 22 months want to know about their child’s vocal development

| Practice Environment |
|----------------------|---------------------|
| Strongly Disagree (%)| Disagree (%)        |
|                      | Neither Agree nor Disagree (%) | Agree (%) | Strongly Agree (%) | Mode (range) |
| 0 (0%)               | 2 (3%)              | 8 (14%) | 30 (52%) | 18 (31%) | 4 (3-5) |

In my opinion, families of children with permanent hearing loss who are younger than 22 months want to know about their child’s first words development

| Practice Environment |
|----------------------|---------------------|
| Strongly Disagree (%)| Disagree (%)        |
|                      | Neither Agree nor Disagree (%) | Agree (%) | Strongly Agree (%) | Mode (range) |
| 0 (0%)               | 0 (0%)              | 1 (2%) | 22 (38%) | 35 (60%) | 5 (3-5) |

### Economic

In my current practice, I have the assessment tools I need to conduct an appropriate assessment of a child with permanent hearing loss’ vocal development before 22 months

| Practice Environment |
|----------------------|---------------------|
| Strongly Disagree (%)| Disagree (%)        |
|                      | Neither Agree nor Disagree (%) | Agree (%) | Strongly Agree (%) | Mode (range) |
| 2 (3%)               | 21 (36%)            | 17 (29%) | 15 (26%) | 3 (5%) | 2 (1-5) |
In my current practice, I have the assessment tools I need to conduct an appropriate assessment of a child with permanent hearing loss’ first words before 22 months

<table>
<thead>
<tr>
<th></th>
<th>0(0%)</th>
<th>8(14%)</th>
<th>9(16%)</th>
<th>24(41%)</th>
<th>17(29%)</th>
<th>4(2-5)</th>
</tr>
</thead>
</table>

In my current practice, I have the time I need to conduct an appropriate assessment of a child with permanent hearing loss’ vocal development before 22 months

<table>
<thead>
<tr>
<th></th>
<th>1(2%)</th>
<th>5(9%)</th>
<th>12(21%)</th>
<th>33(57%)</th>
<th>7(12%)</th>
<th>4(1-5)</th>
</tr>
</thead>
</table>

In my current practice, I have the time I need to conduct an appropriate assessment of a child with permanent hearing loss’ first words before 22 months

<table>
<thead>
<tr>
<th></th>
<th>1(2%)</th>
<th>4(7%)</th>
<th>10(17%)</th>
<th>34(59%)</th>
<th>9(15%)</th>
<th>4(1-5)</th>
</tr>
</thead>
</table>

In my current practice, I have the resources (e.g., access to test forms) I need to conduct an appropriate assessment of a child with permanent hearing loss’ vocal development before 22 months

<table>
<thead>
<tr>
<th></th>
<th>7(12%)</th>
<th>15(26%)</th>
<th>15(26%)</th>
<th>17(29%)</th>
<th>4(7%)</th>
<th>4(1-5)</th>
</tr>
</thead>
</table>

In my current practice, I have the resources (e.g., access to test forms) I need to conduct an appropriate assessment of a child with permanent hearing loss’ first words before 22 months

<table>
<thead>
<tr>
<th></th>
<th>0(0%)</th>
<th>8(15%)</th>
<th>8(14%)</th>
<th>29(50%)</th>
<th>13(22%)</th>
<th>4(2-5)</th>
</tr>
</thead>
</table>

Note: Items in bold meet Cunningham’s (2019) barriers criteria

Given the variability in assessment purpose prioritization, and the very specific sample of SLPs surveyed, we are unable to identify a) which assessment purposes are the representing the assessment purposes most frequently selected by SLPs as important to their clinical practice. Similarly, none of the tests have sufficient evidence that they are sensitive to whether a child who is deaf/hard-of-hearing has acquired new skills or made clinically meaningful progress. In this regard, SLPs’ reports that they do not have the tools they need to complete vocal assessment is consistent with the peer-reviewed literature on available tests. For practicing SLPs, there is currently insufficient evidence to suggest they should adopt or change their current assessment practices.
The assessment purposes prioritized by the SLPs call for inherently different sources of validity evidence, study designs, and statistical analyses than those that are commonly reported in norm-referenced test examiner’s manuals (see Daub, Skarakis-Doyle, et al., 2019 for a discussion). For example, item response theory is a view of test design that is better suited to measuring progress and differential diagnosis but is not currently applied in many tests used by SLPs. In sum, documenting the clinical decisions SLPs plan to make, and those decisions that are the most important to their clinical practice, enables researchers and test developers to design future studies to demonstrate the effectiveness of specific tests in supporting those clinical decisions.

Documenting barriers is similarly expected to inform test and study design to allow tests to fit the clinical contexts in which they will be eventually used. Primarily, barriers related to a lack of available assessment tools and resources (e.g., test forms) for conducting a vocal development assessment. In part, this lack of barriers may reflect the clinical expertise of SLPs who completed our survey. Given the many years of clinical experience as well as advanced training in supporting young CDHH in this particular cohort, we expect that there are likely additional barriers to vocal development assessment across the profession more broadly that were not represented here. Given the present paper’s primary focus of supporting IHP’s decision making, we included many possible barriers and assessment purposes in our survey. As a result, we are underpowered to evaluate whether experience significantly influenced SLPs perception of barriers or their assessment priorities. Future evaluations of how experience influences SLPs assessment priorities would further support the development and adoption of vocal development tests.
5.7.1 Limitations & Future Directions

Because our survey contained assessment purposes developed by the research team, it is possible that the present data do not fully represent all the assessment purposes that SLPs might find useful. A mixed-methods design using qualitative interviewing would have ensured our survey fully captured SLPs’ perspectives. From there, the themes that emerged from the qualitative interview could have guided the development of a survey to explore generalizability. We did attempt, through the use of concept mapping prompts, to collect SLPs’ assessment purposes from their own perspectives, however, we were unable to analyze the statements generated in this survey. Why our prompts failed to elicit statements in the appropriate format highlights a potential challenge with conducting brainstorming sessions remotely, and asynchronously. It is likely that we could have ensured better data quality for analysis if we conducted the brainstorming sessions through teleconference with support from a member of the research team.

However, because of the complex organizational structure of the programs we recruited from, we opted to conduct the brainstorming sessions through survey. Our survey did include questions for asking SLPs to specify any additional decisions that are important to their clinical practice, and the majority indicated that our 15 purposes covered their major purposes. Therefore, we interpret our results as having captured SLPs most important decisions and to be sufficient for developing research priorities. However, by using a survey we were unable to capture the nuanced interpretations that SLPs make using assessment results. Future work using narrative interviews to understand SLPs’ clinical decision making in more depth is expected to also provide rich information about
their assessment needs and to be informative to researchers interested in test development.

With respect to understanding barriers facing vocal development assessment, our use of surveys also limits the conclusions we can draw. In this work, we used a 60% criterion to define statements as barriers out of a need to guide decision making. However, there is no evidence that the barriers identified using this criterion are more, or less, influential in future implementation efforts nor does this criterion allow us to identify for whom the barrier might be greatest. Although it is the case that the majority of SLPs report having the time to complete a vocal assessment, 11% report they do not. Although not in the majority, time may be a major barrier for those SLPs that would make or break the adoption of a new assessment tool. We also do not have information about how barriers relate to practice context – given the regional diversity of SLPs responding to our survey, it may be the case that some SLPs practicing in remote locations experience barriers different from SLPs practicing in urban settings. In this way, the use of a survey design limited our ability to fully understand barriers to assessment, as well as SLPs’ opinions on how to overcome these barriers. Future work using focus groups or interviewing to expand on the barriers documented in our survey data would provide a deeper understanding of SLPs’ practice contexts to researchers looking to develop new vocal development assessments.

Despite these limitations, our work highlights key findings that are of interest to vocal development researchers specifically, and to test developers in speech-language pathology broadly. First, our work reiterates the commercial need for clinically feasible vocal development assessment tools. Despite a lack of appropriate assessment tools,
SLPs in our sample reported that information about vocal development is important to not only their clinical practice, but also to the decision making of other professionals and families of CDHH. SLPs also reported an openness to adopting new tests and using them in their clinical practice. Viewed through the lens of the OMRU, the practice environment and potential adopters, who can dramatically shape the success of implementing a new tool, are supportive of adopting a new tool. In the context of SLPs who work with CDHH in EHDI programs, there do not appear to be attitudinal or organizational barriers to adopting a new test if a suitable one was to become available for them. Future work in vocal development assessment test design could benefit from incorporating SLP feedback about the tool itself in test design to ensure their successful adoption into clinical practice.

Our next step, in partnership with the IHP, is to evaluate SLPs’ initial perception of potential vocal development tests and identify whether SLPs think the tests might be feasible to implement in clinical practice and informative to their clinical decision making. From there, we plan to pilot the tests in clinical practice. During these pilot studies, SLPs will collect data on CDHH’s performance characteristics on tests, and the research team will link these data to audiological variables (e.g., severity of hearing loss, quality of hearing aid amplification) to establish performance characteristics and conduct analyses (e.g., Latent Traits Models) to identify whether the tests are appropriate for the assessment purposes SLPs prioritized here. Finally, at the end of the pilot we will collect data about whether SLPs believed the test was appropriate for the clinical practice, now having experience with the test. We expect that by engaging in this process, we will be
able to support future implementation of a vocal development assessment tool within the
IHP, should a psychometrically appropriate and clinically feasible test be identified.

For the broader research community, our work highlights both the assessment
purposes that SLPs are likely to make using a vocal development assessment and a
methodology for understanding these assessment purposes. Although we cannot draw
conclusions about the extent to which these priorities generalize to the broader SLP
population, we (a) demonstrate the importance of developing tests to fulfill multiple
purposes and (b) present a method that can be used by test developers and researchers to
identify assessment priorities to guide their tool development and validity studies. As
discussed above, these purposes require inherently different validity evidence and our
results provide methodological rationale for researchers designing studies to evaluate the
validity evidence for new tools. To our knowledge, this paper is the first in speech-
language pathology to document the assessment purposes of SLPs with the intention of
incorporating these perspectives into future studies collecting and appraising validity
evidence. We expect that using this approach will allow us to adapt new tools so that
SLPs perceive them as suitable for clinical practice, as well as providing us with a shared
vernacular of assessment purposes with which to discuss these tools with SLPs. Adopting
the perspective of Daub, Skarakis-Doyle, and colleagues (2019), we expect that this
approach will support eventual implementation efforts and clinical uptake.

5.8 Conclusions

SLPs reported numerous vocal development assessment purposes as important to
their clinical practice. The assessment purposes that were prioritized the highest related
to: (a) determining whether a CDHH’s vocal development is within age expectations; (b)
whether the child has made progress; (c) differential diagnosis; and (d) goal setting.

Barriers to vocal development assessment primarily related to a lack of assessment tools.

Future work developing and evaluating vocal development assessments according to these purposes are expected to be beneficial to clinical practice.
References


Chapter 6

6 Introduction

Current best practice recommendations call upon EHDI programs to implement program-wide spoken language outcome monitoring to fulfill a diverse range of purposes (JCIH, 2007, 2013, 2019; Moeller et al., 2013). However, there is no peer-reviewed literature addressing (a) how these assessment purposes might be accomplished and with what tests, (b) whether the methods are implementable within EHDI programs, and (c) whether the procedures result in data that can meaningfully inform stakeholder decision-making. These gaps in the literature have important implications for individual EHDI programs and best-practice recommendations broadly. For individual programs, EHDI programs are faced with the immense task of designing procedures without a framework to guide their process, or a foundation with which to interpret their program outcomes. As this chapter will discuss, measurement is inextricably linked with theory and whether EHDI programs are interpreted as successes or failures rests on the psychometric properties of the tests they use to evaluate their program. For best-practice recommendations, the absence of literature to suggest how (or even whether) spoken language outcome monitoring can be accomplished brings into question the appropriateness of the recommendations themselves. The Joint Committee recommends that spoken language outcome monitoring should be able to answer questions such as: Are services effective? Does a child need additional support? Is the child progressing towards age-appropriate outcomes? These questions are inarguably necessary to answer if EHDI programs are to demonstrate their effectiveness and provide family-centered services. However, the work summarized in this dissertation challenges whether the
recommendations are currently appropriate and, I will argue, highlights limitations in the state of the literature that must be addressed before spoken language outcome monitoring can be considered best-practice.

6.1 Chapter 2

Chapter 2 is a viewpoint article (in press) that argues that the conceptual validity framework described in *Standards* (American Educational Research Association et al., 2014) is a useful framework to ground validity conversations between stakeholders, and offers an appreciable improvement to current testing practices across the profession. Appropriate spoken language outcome monitoring cannot occur if stakeholders are ill-equipped to integrate information from norm-referenced tests into their decision-making. The literature reviewed in this chapter highlights that, within the field of speech-language pathology, the psychometrics underlying tests are broadly misused. Similarly, tests and critical appraisals in speech-language pathology misapply validity concepts. This chapter focused on the relation between decision-making and validity evidence and proposed paths forward to addressing test misuse in speech-language pathology. In relation to spoken language outcome monitoring in EHDI programs, this chapter highlights that the prerequisite assumptions underlying spoken language outcome monitoring (i.e., that tests can be appropriately used to meaningfully inform decision-making) are not met. Significant work is required to address test misuse across the profession and all stakeholders – from SLPs to researchers to test developers – have a role to play in addressing these issues. However, addressing these gaps across the profession does not fulfill the immediate need of EHDI programs like the Ontario IHP to evaluate their outcomes.
Subsequent chapters build from this chapter’s foundation and acknowledge the intersection of decisions and psychometrics. Within our partnership with the Ontario IHP, I was mindful that the data resulting from a spoken language outcome monitoring procedure could be used to make important decisions about programming and resource allocation. Imagining a future where our research team may not be present to support the interpretation of results, my descriptions of a new outcome monitoring procedure would need to recognize the procedure’s inferential limits. I also needed to develop a procedure that was equipped to answer the decisions the IHP planned to make and to aim to protect against inappropriate decision-making.

6.2 Chapter 3

Chapter 3 summarizes the process I engaged in with the IHP, and the research evidence used, to develop recommendations for a spoken language outcome monitoring procedure. The final recommendations included two tiers of testing: (1) program-level outcome monitoring using repeated assessment with omnibus language tests, and (2) outcome monitoring to inform clinical decision-making using tests of spoken language domains known to be at risk in children who are DHH at different ages. Program-level outcome monitoring included assessment every six months until the child turned 3;0 and annually thereafter. I recommended the MBCDI-2 for children younger than 1;6, and the PLS-5 for children between 1;7 and 6;0. Individual vulnerability testing included monitoring key areas of spoken language development known to be at ongoing risk due to inconsistent auditory access (i.e., vocal development and articulation/phonology; vocabulary and grammar; emergent literacy/phonological awareness). This chapter presents, to my knowledge, the first peer-reviewed paper to document the intended
decisions of an EHDI program in spoken language outcome monitoring and to link these decisions to psychometrically appropriate testing methods. This approach recognizes that validity evidence is contextual and allowed me to attempt to build a procedure that protects against inappropriate decision-making. Within the Canadian context, EHDI programs are precariously positioned. Currently, EHDI programs are not a national priority and services for children who are DHH are left to individual provinces to design. This has resulted in incredible inequity across the country, with approximately 35% of the population living in provinces and territories with insufficient services (Canadian Infant Hearing Task Force, 2020). Some provinces such as Ontario and British Columbia have had sufficient EHDI services for decades (Canadian Infant Hearing Task Force, 2014, 2016, 2019, 2020), whereas other provinces (e.g., Saskatchewan, Quebec) have yet to fully implement universal newborn hearing screening, and are missing elements of EHDI programing. Other provinces (e.g., New Brunswick) have seen declines in support and service provision, indicating that merely establishing adequate EHDI services is not enough. Rather, efforts need to be ongoing to ensure that adequate services are sustained.

This context provides an enormous opportunity for data from programs such as the Ontario IHP to advocate for equitable access to EHDI programming. Therefore, careful consideration of how the spoken language outcome monitoring procedure would be designed was required to protect against misinterpretation of outcomes that could jeopardize the IHP’s program and demonstrate the benefit of services on national platforms.

The major contributions of this chapter, therefore, are two-fold. First, this chapter provides a public record of the types of decisions that a spoken language outcome
monitoring procedure was designed to answer, and recognition of the inferential limits of that procedure. In documenting, for instance, that some EHDI programs observe group level performance at the 30th percentile on the PLS-5, this manuscript can be used as evidence that the 30th percentile is an acceptable outcome. If misunderstood, as is a risk outlined in Chapter 2, PLS-5 data at the 30th percentile could be used to erroneously conclude that IHP services are ineffective and used as evidence to defund the program.

Second, this chapter provides a preliminary framework for developing a spoken language outcome monitoring procedure that can be tailored to other EHDI programs. I do not assert that the procedure described in this chapter is a gold-standard, but it is a beginning that can be refined through engagement with researchers and stakeholders.

6.3 Chapter 4

Where Chapter 3 documented the process I used for designing the spoken language outcome monitoring procedure, Chapter 4 reports on a preliminary evaluation to determine if the procedure can be (a) used to support the Ontario IHP’s intended decisions and (b) implemented in clinical practice. Data from this chapter were collected over the course of a 12-month pilot where SLPs working in the IHP administered both the program-level and individual vulnerability testing procedures and provided feedback on their experience at the end of the pilot. Previous efforts to evaluate spoken language outcome data maintained in the IHP database (i.e., Daub, 2016; Daub et al., 2017) highlighted that there were significant data limitations in implementation of the previous outcome monitoring procedure (i.e., administration of the PLS-4 every six months). That is, many children were missing PLS-4 data (for unknown reasons) and, even when data were present, there were often inconsistencies in data entry or scoring. Reasons for
inconsistencies could not always be clearly linked to SLP administration, scoring, or data entry errors. These issues around missing and incomplete data limited the questions I was able to answer for the IHP in previous analyses, and certainly did not encapsulate the scope of questions the IHP aimed to answer on the basis of Joint Committee recommendations. In order to justify the use of a new procedure, it was necessary to collect evidence that SLPs could (and would) consistently and accurately implement the procedure, and that the data could be used to answer the IHP’s questions.

Results from the pilot study reported in Chapter 4 provide early evidence that the program-level procedure could be used to meaningfully inform the IHP’s decisions. I was able to identify some suggestive evidence that the PLS-5 Growth Scale Values (but not Standard Scores) could be sensitive to change over time (consistent with Daub et al., 2017), but neither the MBCDI-2 nor PLS-5 conformed to predictions when it came to identifying whether scores were influenced by predictors of spoken language. SLPs also reported that, although they were confident in their ability to implement the procedure, they were concerned about the time the procedure would take away from other areas of their clinical practice. In this pilot study, time to implement the program-level procedure was confounded with also administering the individual vulnerability testing procedure, given that 10 of the 11 regions were simultaneously piloting both procedures.

Results for the individual vulnerability testing procedure were even less clear. There was insufficient data to perform the planned analyses, which were intended to relate performance on individual tests to one another as well as to the program-level procedure. My original intention was to develop a structural equation model relating the program-level and individual vulnerability testing procedures, with the goal of supporting
development of intervention pathways or decision-making guidelines for (re-)initiating SLP intervention. Within the IHP, children who are DHH and learning a spoken language are not always receiving active supports from SLPs. At various times, the decision might be made by the child’s family and SLP that they are not currently concerned about the child’s spoken language development and so the child might be discharged. However, because inconsistent auditory access (Moeller & Tomblin, 2015) is ongoing for children who are DHH, they may begin to struggle in some domains of spoken language despite earlier performance within normal limits (Moeller et al., 2007b). I thought that the individual vulnerability testing procedure would serve as a mechanism to monitor these key areas of spoken language so that SLPs could reinstate active intervention if concerns began to emerge. Although the program-level procedure does not preclude SLPs from monitoring these areas, consistent individual vulnerability testing throughout the IHP was expected to be able to support equitable service across the province. For instance, it could be the case that some SLPs in better resourced areas would have the time and resources to monitor areas more closely than other regions. I also thought that these data could be used to develop new IHP guidelines to support consistent decision-making and service allocation across regions. Even in regions where SLPs may have the resources to monitor individual vulnerability testing of their own volition, they may not have the ability – practicing within a publicly-funded health system – to reinstate services. Hypothetically, the evidence collected by an individual vulnerability testing procedure could be used to advocate at a systems level for the need for ongoing, more intensive intervention for those children who show signs of needing additional services.
Despite these intentions, the individual vulnerability data collected in the pilot study were insufficient to fulfill any of my planned purposes. In addition, as was the case with the program-level procedure, SLPs were concerned about the amount of time involved in implementing the procedure. Given that there was more evidence that the program-level procedure could support decision-making, and because it was possible that implementing the individual vulnerability procedure could jeopardize the implementation of the program-level procedure, my final recommendation for the IHP was to formally adopt the program-level procedure and not the individual testing procedure. Interestingly, however, more SLPs reported that the individual vulnerability testing procedure was more informative to their clinical practice than the program-level procedure. Because of this, and because of the theoretical importance of individual vulnerability testing, I recommended that the IHP allocate funding to providing SLPs with the test materials to conduct individual vulnerability testing when they deemed it appropriate.

6.4 Chapter 5

Part of the original recommendations for individual vulnerability testing included monitoring of children’s early vocal development, encompassing infraphonological development to canonical babble, early consonant inventories, and syllable structures. Delayed onset of canonical babble (Oller, 2000) as well as restricted consonant inventories and syllable structures (Moeller et al., 2007a) have been demonstrated to predict children who are at increased risk of overall spoken language delays. Theoretically, early vocal development is one of the earliest markers that a SLP could use to identify whether a child needs additional intervention, or whether there are other factors influencing their language development above and beyond the child’s hearing
impairment (e.g., a language learning disorder more broadly, a comorbid motor speech disorder) that requires intervention. As a part of the individual vulnerability testing procedure, vocal development assessments could be used to identify children who may need additional intervention services early in development. Based on the review described in Chapter 3, I identified that there were currently no vocal development tests that had appropriate psychometric evidence to include in the individual vulnerability testing procedure. However, there are several tests currently in development that may be appropriate to recommend with additional evidence.

Grounded in Standard’s (2014) conceptual validity framework (Chapter 2), Chapter 5 reported the first peer-reviewed study documenting SLPs’ vocal development assessment purposes (or, intended decisions – Chapter 2) and barriers to assessing vocal development. Results from this study were expected to inform future partnerships with the Ontario IHP to collect validity evidence addressing the assessment priorities most consistently prioritized by SLPs. Methodologically, this chapter is a departure from traditional approaches to validity studies, which focus on collecting evidence for a narrow range of purposes, often aligned with psychometric checklists (Daub et al., 2019; Daub et al., in press). As discussed in Chapter 2, these checklists do not contextualize evidence within decision-making. In preparing this study, I expected that linking validity evidence with the assessment purposes identified as prioritized in Chapter 5 will, in future work, support SLPs in appropriate decision-making based on assessment results. Additionally, I expected that documenting barriers would inform the future selection of vocal development tests to identify those candidate tests with administration properties most likely to overcome barriers (e.g., prioritizing shorter tests). Knowing about both
SLPs’ assessment purposes and the barriers they encounter would allow programs to select a test that is both psychometrically equipped and the most clinically feasible. In effect, I expect that the information gathered in Chapter 5 will allow the IHP (and other EHDI programs) to design procedures most likely to be successful in future pilot studies, such as the work described in Chapter 4.

Results from this chapter highlight that SLPs would use a hypothetical vocal development test for a number of purposes, many of which are not supported by commonly reported validity evidence in norm-referenced tests (e.g., differential diagnosis). This chapter also highlighted that if an appropriate vocal development test were identified, it could be appropriately implemented in clinical practice. The questions relating to barriers were based on surveys used by Moodie and colleagues (2011) to inform the development and implementation of the UWO PedAMP (an auditory outcome monitoring protocol currently adopted by the IHP; Bagatto et al., 2011). Moodie’s original surveys were grounded in the diffusion of innovations theory (Rogers et al., 2003), which offers theoretical explanations for why (a) new innovations are adopted and (b) adoption spreads throughout a group (in our case, SLPs). The majority of SLPs reported that they, their colleagues, and clients would find the results of vocal development assessments meaningful. The majority of SLPs also indicated that they had the skills, although not the means, to administer vocal development assessments. Overall, this chapter indicates the clinical need for new vocal development tests and the purposes those tests ought to fulfill (in an EHDI context). Future work collecting validity evidence on new tools and incorporating SLPs’ perspectives about the tools early in development could facilitate eventual implementation success (Daub et al., 2019).
6.5 Implications: Supporting the Ontario IHP

One of this dissertation’s main contributions is the development of a spoken language outcome monitoring procedure that can be implemented within the Ontario IHP. Despite the complexities underlying test use (Chapter 2) and assessment for multiple purposes (Chapter 3), I was able to develop a procedure (Chapter 3) with data that may support IHP decision-making (Chapter 4). I also report preliminary evidence that children in the IHP may achieve spoken language outcomes that not only meet, but surpass, the outcomes previously reported for children in EHDI programs. Data from children without additional factors influencing outcomes had higher than expected outcomes (Chapter 4). If this pattern holds in future analyses of IHP outcomes, this would indicate that the IHP is not only meeting but surpassing our initial expectations and lend support to (a) continuing the provision of the IHP in Ontario (b) advocacy efforts for EHDI programs nationwide, and c) demonstrating the effectiveness of current IHP policies and procedures so that they may be adopted (or modified) by other EHDI programs.

Despite these important contributions, the procedure has some limitations. The individual vulnerability testing procedure itself was incomplete (i.e., missing a vocal development test; Chapter 5) at the time of the pilot and it is likely that adding another test will increase the amount of time the procedure takes (and thus increase SLP concerns in this regard). The procedures are also only applicable for children with sufficient levels of English language exposure and do not clearly operationalize expected outcomes for children with additional diagnoses (who account for approximately 30 – 40% of children who are DHH; Cupples et al., 2014). These limitations have important equity implications. While the results from Chapter 4 suggest that the program-level procedure
can be reasonably implemented to produce usable data, the children for whom the procedure is appropriate is not representative of the broader Ontario population. As a first effort to build a spoken language outcome monitoring procedure, I needed to limit the procedure to include only English spoken language tests largely given my own language proficiency, the languages represented in the peer-reviewed literature, and a general lack of norm-referenced tests for languages other than English. This is insufficient. An EHDI program cannot equitably evaluate its service provision if data from large groups of children are not represented. Only 77.6% of Ontarians report speaking only English at home (Statistics Canada, 2018) and this varies by province with the lowest percentage of English monolingual speakers living in Quebec (9.7%; Statistics Canada, 2018). In Ontario, excluding the 22% of the population speaking a language other than English at home from program evaluations fails to identify whether services are adequate for all children in the province, prevents the IHP from identifying whether inequities exist within the program, and prevents them from developing solutions to overcome these inequities.

Future work is needed to identify solutions so that the IHP can accurately capture outcomes for all children in the program. Given a lack of norm-referenced tests for many languages, this will likely necessitate a different monitoring approach. Possible solutions could involve using communicative participation outcome measurement tools that are already in use by other publicly funded speech-language pathology services in Ontario and for which there are many translations (i.e., the Focus On Communication Outcomes Under 6 used in the Ontario Preschool Speech and Language program; Cunningham et al., 2018; Thomas-Stonell et al., 2010). Another possibility could include co-producing
benchmark checklists with SLPs and families who access IHP services that focus more specifically on spoken language milestones (e.g., syllable complexity, grammatical development) rather than communicative participation.

The Ontario IHP ultimately adopted the new program-level procedure in 2020 for province-wide implementation, and the individual vulnerability testing procedure is under consideration for modified implementation. It remains to be seen whether the procedure can be sustainably implemented and results in data that can meaningfully inform decisions once data are entered into the provincial database and analyzed. The IHP has agreed to monitor implementation on an ongoing basis through audits of its database in order to identify whether data are entered consistently and accurately, and whether the data are analyzable in the ways in which they were intended. Unfortunately, due to the COVID-19 pandemic, questions surrounding whether the procedure is possible to implement are magnified. In Ontario, face-to-face service provision was halted at the time the IHP planned to implement the new program-level procedures and in-person service currently remains limited to only those situations where virtual, or postponing, services cannot occur. While the PLS-5 was originally preferred for program-level outcome monitoring even at the youngest ages, it is interesting that the MBCDI-2 has become most useful during the pandemic because it can be administered remotely (whereas the PLS-5 is difficult to reliably administer virtually). The timing of resuming in-person services, and whether timelines will be similar across regions of the province, is currently unknown, but these factors are likely to influence the successful implementation of the new spoken language outcome monitoring procedures.
6.6  Implications: Evaluating EHDI best practice recommendations for outcome monitoring

In addition to providing practical recommendations for the Ontario IHP, this dissertation initiates a conversation surrounding the appropriateness of recommendations for routine spoken language outcome monitoring. This dissertation was centered around the simplest case of spoken language outcome monitoring: monitoring with English, norm-referenced tests for the primary purpose of program evaluation, but from which additional information for clinical purposes could be drawn. Despite this focus, there remain significant barriers to accomplishing spoken language outcome monitoring in a sustainable and equitable away. Put simply, spoken language outcome monitoring is immensely complicated from psychometric and implementation perspectives and current recommendations do not sufficiently address these considerations.

6.6.1  Psychometric limitations of spoken language outcome monitoring

At its core, this dissertation is concerned with answering whether EHDI programs can validly make decisions on the basis of spoken language outcome monitoring data. In attempting to answer this question, new challenges emerged in every chapter. Chapter 2 highlights that there are significant gaps in the ways that test evidence is used across the profession (including the design and critical appraisal of individual tests themselves) that are not unique to EHDI outcome monitoring. Test misuse, and a misapplication of validity concepts, is pervasive in all areas of the profession from clinical use (Kerr et al., 2004; McCauley & Swisher, 1984), to basic scientific research (Nitido & Plante, 2020), to test design (Daub et al., 2019), to critical appraisal (Daub et al., in press). As Chapter 3 highlights, spoken language outcome monitoring cannot be accomplished – for the
purposes it is intended – with a singular test. Complex applications of psychometrics, such as integrating *multiple* tests and scores (each for a unique purpose) cannot occur if tests are misused at their simplest application. Chapter 3 attempts to overcome these fundamental issues with test use by making explicit the link between decisions and validity evidence and delineating the inferential limits of the proposed procedures.

Chapter 4 attempts to collect validity evidence for the use of the procedure in clinical practice. In addition to concerns about the feasibility of implementing both the program-level and the individual vulnerability testing procedures, Chapter 4 uniquely highlights the inextricable link between measurement and theory. With regard to the (unexpected) null finding of the influence of severity of hearing loss (better-ear pure tone average) and spoken language outcomes, it is unclear whether the null effect is an artefact of our procedure or our sample, or a theoretical truism of the data. In previous work (see Chapter 3), the performance of children who are DHH on the PLS-4 was regularly predicted by severity of hearing loss, which was one of the reasons I selected the PLS-5 to accomplish program-level monitoring. Theoretically, children who are DHH and who receive adequate services should perform within normal limits on an omnibus language test. As a group, this is what I observed in Chapter 4 – children without additional factors influencing their performance mirrored a normative distribution with a group mean at the 50th (not 30th) percentile. However, we cannot infer typical development based on our data. It is possible that the PLS-5 was not sensitive enough to detect spoken language domains where children struggled (thus the motivation for the individual vulnerability testing).
Disentangling whether the results observed in Chapter 4 are a measurement artefact or evidence that the children in the dataset were (as a group) truly performing within normal limits is problematic. EHDI programs must make the determination that their procedure is informative with no control group. Because there is evidence that children who are DHH who receive EHDI services have improved outcomes over those who do not (e.g., Ching et al., 2013; Tomblin et al., 2015) and because hearing loss can impose lifelong consequences when intervention is delayed (e.g., Nelson et al., 2008; Patel & Feldman, 2011), audiologists and speech-language pathologists do not have clinical equipoise. Therefore, EHDI programs cannot deny some groups of children intervention to determine whether an outcome monitoring procedure is truly sensitive to the differences between adequate and inadequate levels of intervention.

Chapters 3 and 5 also highlight the unique challenges EHDI programs face in trying to design their own procedures. In order to carefully inform intervention and determine whether a child is progressing towards age-appropriate outcomes, programs need tools that are (a) sensitive to change over time and (b) predictive of future outcomes. Quite simply, there is no singular tool that is sufficient to accomplish these purposes. Insufficient evidence that tools necessary to accomplish these purposes exist, and this is particularly pronounced in the case of assessing vocal development assessment. One approach to overcome the psychometric limitations of using a single test to measure spoken language outcomes could be to use assessment batteries (e.g., Tomblin et al., 2015) to derive composite scores of overall spoken language performance. Under classical test theory paradigms, we could increase the precision with which we estimate children’s ability by increasing the number of observations (or test scores) to attempt to
reduce measurement error. However, increasing measurement precision also has the potential to increase the clinical burden (i.e., time) of implementing a procedure.

6.6.2 Implementation limitations of spoken language outcome monitoring

Just as psychometrics are inextricably linked with theory, so too are psychometrics and test use. Data collected from a spoken language outcome monitoring procedure cannot be validly used to inform decisions if the tests are not psychometrically appropriate for their purposes. But increasing measurement precision can increase the burden associated with implementing the procedures, as SLPs reported in Chapter 4. If procedures are not appropriately implemented, the resulting data may not be psychometrically equipped to validly inform decision making.

There is a vast body of literature documenting that (a) research evidence rarely influences clinical practice and (b) psychological (e.g., Atkins et al., 2017) and sociological (e.g., Rogers, 2003) factors influence whether research is used. The pilot study I conducted in Chapter 4 was intended to identify whether SLPs believed the spoken language outcome monitoring procedures were feasible to implement, or whether modifications were needed to either the procedures, the online learning module used to present the procedures, or their employment context (e.g., additional time allocated to assessment) to support implementation. Of key importance is that SLPs completed the surveys after having used the procedure for 12 months. Beliefs about an evidence-informed innovation (our spoken language outcome monitoring procedure) influence implementation, but beliefs change over time and experience with an innovation (Rogers, 2003). Previous work with the IHP identified that, after reviewing the procedure and
online learning modules, SLPs rated the procedure, the way it was designed, its potential clinical value, and the online learning module favorably, with an intention to implement the procedure (Cunningham et al., 2019). Pre-post analyses of the survey data were beyond the scope of Chapter 4, but overall, more SLPs reported feeling neutral (as opposed to positively) about the strength of evidence underlying the program-level procedure, and its potential value, than they did in Cunningham et al.’s original evaluation. In both instances, many SLPs reported that the time involved in both procedures was inappropriate for clinical practice. However, fewer SLPs reported that they agreed or strongly agreed with the tests included in the program-level procedure in Chapter 4 than they did initially. Whether these changes in perception will continue as implementation spans the course of years, and whether these perceptions influence implementation, is unknown but has importance for the sustainability of the procedure.

The influence of implementation on data usability was observed in Chapter 4 (given the large amount of missing data, and questions about sample specificity) and in previous evaluations of the IHP’s program (Daub et al., 2017). With full implementation across the IHP, it will be necessary to monitor through regular audits of the provincial database whether assessments are completed routinely and scored accurately. Although Chapter 4 reports preliminary evidence that the program-level procedure is feasible to implement, contrary evidence may emerge over the next few years.

6.6.3 Ethical considerations underlying spoken language outcome monitoring

This dissertation has focused on the interplay between assessment decisions, psychometrics, and implementation in the context of spoken language outcome
monitoring in EHDI programs. How these three elements interact influences whether decisions can be validly made, by different stakeholders, on the basis of spoken language outcome monitoring data. Equally important yet currently unanswered questions are (a) whether the decisions made based on spoken language outcome monitoring result in meaningful action and (b) whether the actions result in improved EHDI services for all children who are DHH.

Assessment is not a risk-free activity. Conducting assessments costs time and money for EHDI programs, SLPs, and families that could be allocated elsewhere. So too does entering, maintaining, and analyzing the data. But the greater risks relate to the harm that can be imparted on individual children, their families, children who are DHH, and marginalized communities. When inappropriately collected or misinterpreted, data can be used to create (or reinforce) systemic barriers to healthcare (Messick, 1993) by drawing erroneous conclusions. For instance, stakeholders could use (hypothetical) poorer outcomes for marginalized communities to conclude there is a lack of evidence for treatment benefit or as an impetus to better understand social determinants of health. The issue around misinterpretation is particularly pronounced in the EHDI context when stakeholders look to understand whether children have made change over time. If stakeholders use standard scores without understanding their psychometric properties, they might be inclined to interpret a 0-point change in standard scores as evidence of “no growth”. This could be used to argue that children are not progressing (and therefore need more intensive intervention) or that EHDI intervention itself is ineffective. However, because standard scores measure relative standing rather than absolute ability, a 0-point change is more reasonably evidence that a child maintained their standing.
relative to same-aged peers between assessments. In order to do so, they would have necessarily had to have acquired new skills at a rate commensurate with their same-aged peers (Daub et al., 2017).

Given the risks associated with spoken language outcome monitoring, particularly to children who are DHH, their families, and marginalized communities, EHDI programs must ensure that spoken language outcome monitoring can be appropriately implemented and meaningfully inform stakeholders decisions.

6.6.4 Overall Implications for EHDI best practice recommendations

Although this dissertation has offered a (tentative) path for the Ontario IHP and other EHDI programs to accomplish spoken language outcome monitoring, it has highlighted – for the first time in the peer-reviewed literature – the psychometric and implementation limitations of doing so. Fundamentally, I believe that the questions that best practice recommendations seek to answer using spoken language outcome monitoring are important ones. Programs and clinicians need to know that their services are effective, and parents need to know what supports their children need and whether modifications to intervention are needed. In publicly funded systems, taxpayers need to know that their tax dollars are allocated to effective services. Equitable service provision across the country needs data to compare between EHDI programs. But if stakeholders do not grapple with the challenges highlighted in this dissertation, EHDI programs who (acting in good faith) attempt to implement best practice recommendations risk doing more harm than good.
6.7 Future directions

There is considerable work to be done, from the perspective of psychometrics and implementation science, to reduce the risk associated with spoken language outcome monitoring. Evidence that procedures can meaningfully answer stakeholder questions, using the data that are collected within EHDI programs over time, is needed. The immediate first step is to return to fundamental considerations of decisions and evidence. As a field, we need to clearly articulate which decisions EHDI programs need to make, how they should be prioritized if not all decisions can be feasibly accomplished, and the sources of evidence that will compel stakeholders. The Ontario IHP prioritized the consistent implementation of the program-level procedure to answer whether children were performing within age-expectations and making progress over time. This will allow SLPs the flexibility to make their own determination whether intervention is needed and when, for individual children using results from their clinical assessments rather than a standard battery. Other programs may not have the resources to appropriately implement a regular outcome monitoring procedure using norm-referenced assessments every six months or annually. In these cases, programs might reasonably elect to evaluate spoken language outcomes with a norm-referenced test only at program discharge (to ensure whether the child achieved age-appropriate outcomes) and monitor regularly using benchmarks or checklists. Whether these approaches may be considered best practice are contextual to a program’s individual circumstances. With decisions identified, stakeholders in the spoken language outcomes of children who are DHH can begin the important task of developing, evaluating, implementing, monitoring, and sustaining spoken language outcome monitoring procedures.
6.8 Conclusions

Spoken language outcome monitoring is not easily accomplished and poses risks to children who are DHH if inappropriately applied. This dissertation contributed a candidate procedure and preliminary data of its utility, to support a Canadian EHDI program, Ontario’s Infant Hearing Program, in accomplishing spoken language outcome monitoring. This dissertation also highlighted the challenges underlying spoken language outcome monitoring and the significant work that is urgently needed before spoken language outcome monitoring can be considered best practice for EHDI programs worldwide.
6.9 References


Appendices

Appendix 1: CINAHL Search Strategy

#1 (MH "Outcome Assessment") OR (MH "Outcomes (Health Care)") OR (MH "Treatment Outcomes")
#2 (MH "Child, Disabled") OR (MH "Child, Preschool") OR (MH "Child Health") OR (MH "Child Development Disorders")
#3 (MH "Hearing Loss, Functional") OR (MH "Hearing Loss, Partial") OR (MH "Hearing Loss, Sensorineural") OR (MH "Hearing Loss, Conductive") OR (MH "Hearing Disorders") OR (MH "Deafness")
#4 (MH "Language") OR (MH "Speech and Language Assessment") OR (MH "Rehabilitation, Speech and Language") OR (MH "Language Disorders")
#5 (MH "Outcome Assessment") OR (MH "Outcomes (Health Care)") OR (MH "Treatment Outcomes")
#6 (MH "Child, Disabled") OR (MH "Child, Preschool") OR (MH "Child Health") OR (MH "Child Development Disorders")
#7 (MH "Hearing Loss, Functional") OR (MH "Hearing Loss, Partial") OR (MH "Hearing Loss, Sensorineural") OR (MH "Hearing Loss, Conductive") OR (MH "Hearing Disorders") OR (MH "Deafness")
#8 (MH "Language") OR (MH "Speech and Language Assessment") OR (MH "Rehabilitation, Speech and Language") OR (MH "Language Disorders")
#9 S5 AND S6 AND S7 AND S8
#10 (MH "Clinical Assessment Tools") OR (MH "Speech and Language Assessment") OR (MH "Outcome Assessment") OR (MH "Functional Assessment")
#11 (MH "Instrument Validation")
#12 (MH "Clinical Assessment Tools")
#13 (MH "Language Tests")
#14 ((MH "Language Tests") AND (S1 OR S10 OR S11 OR S12 OR S13))
#15 (((MH "Language Tests") AND (S1 OR S10 OR S11 OR S12 OR S13)) AND (S5 OR S14))
#16 (((MH "Language Tests") AND (S1 OR S10 OR S11 OR S12 OR S13)) AND (S5 OR S14)) AND (S6 AND S7 AND S8 AND S15)
#17 (MH "Measurement Issues and Assessments")
#18 ((MH "Measurement Issues and Assessments") AND (S1 OR S5 OR S10 OR S11 OR S12 OR S13 OR S17))
#19 ((MH "Measurement Issues and Assessments") OR S1 OR S5 OR S10 OR S11 OR S12 OR S13 OR S17)
#20 ((MH "Measurement Issues and Assessments" OR S1 OR S5 OR S10 OR S11 OR S12 OR S13 OR S17)) AND (S3 AND S4 AND S6 AND S19)
#21 ((MH "Measurement Issues and Assessments" OR S1 OR S5 OR S10 OR S11 OR S12 OR S13 OR S17)) AND (S3 AND S4 AND S6 AND S19)
#22 (MH "Infant") OR (MH "Infant Development")
#23 (MH "Early Childhood Intervention")
#24 ((MH "Early Childhood Intervention") OR (S6 OR S22 OR S23)
#25 ((MH "Early Childhood Intervention") OR (S6 OR S22 OR S23)) AND (S7 AND S8 AND S19 AND S24)

Appendix 2: Data extraction

Supplemental materials describing the 36 studies can be found on Open Sciences Framework [https://osf.io/ncm23/?view_only=1455217c19c44e3881e4628ed252fe3a](https://osf.io/ncm23/?view_only=1455217c19c44e3881e4628ed252fe3a)

Details such as study authors, tests used, sample characteristics, and study purposes are laid out in an easy-to-read table. We also list whether the authors included composite scores, made group comparisons, noted informal differences, and evaluated change over time. Finally, we noted if the study had statistically significant results or if they included other analyses.
Appendix 3: Children with one PLS-5 assessment hearing aid Speech Intelligibility Index (SII) in conversational speech, compared to Moodie et al., 2017 normative data
Appendix 4: Children with one PLS-5 assessment hearing aid Speech Intelligibility Index (SII) in quiet speech, compared to Moodie et al., 2017 normative data.
Appendix 5: Children with one MBCDI-2 assessment hearing aid Speech Intelligibility Index (SII) in conversational speech, compared to Moodie et al., 2017 normative data
Appendix 6: Children with one MBCDI-2 assessment hearing aid Speech Intelligibility Index (SII) in quiet speech, compared to Moodie et al., 2017 normative data
Appendix 7: Tier 1 post-pilot survey

Ontario Infant Hearing Program End of Pilot Survey - Program Level Outcome Monitoring

Thank you for participating in the pilot study to implement new spoken language outcome monitoring procedures in Ontario’s Infant Hearing Program (IHP). We appreciate the time and commitment of everyone who supported this initiative and for the valuable data and feedback you have provided thus far.

We are now asking for your feedback after having trialed the new procedures for one year. In this survey, we wish to learn about your experiences with the new Program-Level Outcome Monitoring (MacArthur-Bates Communicative Development Inventories (CDI) Words and Gestures/Preschool Language Scales, Fifth Edition (PLS-5)) procedures.

Your input is extremely valued and will be used to inform decisions about and procedures for province-wide implementation of a new Program-Level Outcome Monitoring process in the IHP. Thank you for sharing your honest perspectives from the front-lines!

Note: Please respond to items in this survey considering only your experiences with the Program Level Outcome Monitoring procedures (i.e., CDI and PLS-5). A separate feedback survey is required for those who participated in the Individual Vulnerability Testing subpilot.

It is essential that we be able to link your responses on this survey with the three surveys you completed at the beginning of the pilot. To ensure that you remain anonymous on all of these surveys, we asked you to create an anonymous username at the beginning of the pilot to use for all of the surveys.

As a reminder, this username was required to be at least 8 characters, but was otherwise allowed to contain any typet of characters. We asked you to record your anonymous username somewhere safe and easy for you to access so that you would be able to enter it on future surveys.

Please type that same anonymous username in the box:

Roughly how many IHP children did you attempt to assess using the new Program Level Outcome Monitoring Procedures (i.e., MCDI and PLS-5) during the pilot?

- None
- 1 to 5
- 6 to 10
- more than 10
To begin, please answer the following two general questions about your experiences with the new Program Level Outcome Monitoring procedures over the past year.

<table>
<thead>
<tr>
<th>The IHP’s new Program Level Outcome Monitoring procedure was useful for my clinical practice.</th>
<th>Not applicable</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>I was able to consistently implement the new Program Level Outcome Monitoring recommendations in my practice.</th>
<th>Not applicable</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>Response Options</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The length of time it took to administer the recommended Program Level Outcome Monitoring tests was appropriate for incorporation into routine clinical practice.</td>
<td>〇 Not applicable 〇 Strongly disagree 〇 Disagree 〇 Neutral 〇 Agree 〇 Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td>〇 Not applicable 〇 Strongly disagree 〇 Disagree 〇 Neutral 〇 Agree 〇 Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>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.</td>
<td>〇 Not applicable 〇 Strongly disagree 〇 Disagree 〇 Neutral 〇 Agree 〇 Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The time it took to do the recommended Program Level Outcome Monitoring and reporting did NOT negatively impact other areas of my practice.</td>
<td>〇 Not applicable 〇 Strongly disagree 〇 Disagree 〇 Neutral 〇 Agree 〇 Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The environment in which I worked made it difficult for me to implement the recommended Program Level Outcome Monitoring.</td>
<td>〇 Not applicable 〇 Strongly disagree 〇 Disagree 〇 Neutral 〇 Agree 〇 Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had the supplies I needed (e.g., test forms) to implement the new Program Level Outcome Monitoring.</td>
<td>〇 Not applicable 〇 Strongly disagree 〇 Disagree 〇 Neutral 〇 Agree 〇 Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When I had a question about the Program Level Outcome Monitoring procedures, I consulted with my colleagues.</td>
<td>〇 Not applicable 〇 Strongly disagree 〇 Disagree 〇 Neutral 〇 Agree 〇 Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When I had a question about the Program Level Outcome Monitoring procedures, I consulted with my managers/administrators.</td>
<td>〇 Not applicable 〇 Strongly disagree 〇 Disagree 〇 Neutral 〇 Agree 〇 Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>Options</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When I have a question about the Program Level Outcome Monitoring procedures, I consulted the &quot;Pilot Implementation Q&amp;A&quot; section of Western's OWL site</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had the resources I needed (e.g., administrative support for scheduling, data entry) to do the new Program Level Outcome Monitoring procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had permission from my manager to take the time I needed to complete Program Level outcome monitoring procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Getting timely feedback from experts (i.e., the research team at Western University) helped me to implement the new Program Level Outcome Monitoring procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The SLPs I worked with were excited about the new Program Level Outcome Monitoring procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managers/administrators I worked with were supportive of the new Program Level Outcome Monitoring procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The parents I worked with were interested in the results of the new Program Level Outcome Monitoring procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The task of completing the MCDI was not too difficult for parents (respondents) to perform.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The task of completing the MCDI was not too time consuming for parents (respondents) to perform.  

- Not applicable  
- Strongly disagree  
- Disagree  
- Neutral  
- Agree  
- Strongly agree

Please provide any comments related to the above questions regarding how your clinical practice environment impacted the implementation of the new Program-level outcome monitoring procedures over the past year.

[Blank line]
Please select one answer for each of the following questions about how your clinical knowledge, skills, and beliefs affected your implementation of program level outcome monitoring procedures over the past year.

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over the past year I felt I had the clinical skills required to implement the new Program Level Outcome Monitoring procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>I am familiar with the administration of the MacArthur-Bates Communicative Development Inventories Words &amp; Gestures (MCDI).</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>I was able to accurately score and use the norms tables for the MacArthur-Bates Communicative Development Inventories Words &amp; Gestures (MCDI).</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>I am familiar with the administration of the Preschool Language Scales-5th Edition.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>I was able to accurately score and use the norms tables for the Preschool Language Scales-5th Edition.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>The new Program Level Outcome Monitoring procedures have helped me with my clinical decision-making.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>The new Program Level Outcome Monitoring procedures have helped parents with their decision-making.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>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.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
</tbody>
</table>
Repeat administration of the Program Level Outcome Monitoring tools to the same child was useful for my own clinical practice.

- Not applicable
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Please provide any comments related to the above questions regarding how your knowledge, skills, or beliefs impacted the way Program-level outcome monitoring procedures were implemented over the past year.

______________________________

23/05/2021 5:00pm
www.projectredcap.org
Please select one answer for each of the following questions about how the development or content of the program level outcome monitoring procedures impacted how they were implemented over the past year.

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>The new Program Level Outcome Monitoring procedures were similar to the previous outcome monitoring procedures for the IHP.</td>
<td>Not applicable&lt;br&gt;Strongly disagree&lt;br&gt;Disagree&lt;br&gt;Neutral&lt;br&gt;Agree&lt;br&gt;Strongly agree</td>
</tr>
<tr>
<td>The new Program Level Outcome Monitoring procedures were an improvement over the current procedure.</td>
<td>Not applicable&lt;br&gt;Strongly disagree&lt;br&gt;Disagree&lt;br&gt;Neutral&lt;br&gt;Agree&lt;br&gt;Strongly agree</td>
</tr>
<tr>
<td>I found the MacArthur-Bates Communicative Development Inventories Words &amp; Gestures to be a high-quality clinical outcome evaluation tool.</td>
<td>Not applicable&lt;br&gt;Strongly disagree&lt;br&gt;Disagree&lt;br&gt;Neutral&lt;br&gt;Agree&lt;br&gt;Strongly agree</td>
</tr>
<tr>
<td>I found the MacArthur-Bates Communicative Development Inventories Words &amp; Gestures to be a valid and reliable tool for preschoolers with permanent hearing loss.</td>
<td>Not applicable&lt;br&gt;Strongly disagree&lt;br&gt;Disagree&lt;br&gt;Neutral&lt;br&gt;Agree&lt;br&gt;Strongly agree</td>
</tr>
<tr>
<td>I felt the MacArthur-Bates Communicative Development Inventories Words &amp; Gestures was the right choice for evaluating spoken language outcomes for the IHP's youngest children.</td>
<td>Not applicable&lt;br&gt;Strongly disagree&lt;br&gt;Disagree&lt;br&gt;Neutral&lt;br&gt;Agree&lt;br&gt;Strongly agree</td>
</tr>
<tr>
<td>I found the Preschool Language Scales-5th Edition to be a high-quality clinical outcome evaluation tool.</td>
<td>Not applicable&lt;br&gt;Strongly disagree&lt;br&gt;Disagree&lt;br&gt;Neutral&lt;br&gt;Agree&lt;br&gt;Strongly agree</td>
</tr>
<tr>
<td>I found the Preschool Language Scales-5th Edition to be a valid and reliable tool for preschoolers with permanent hearing loss.</td>
<td>Not applicable&lt;br&gt;Strongly disagree&lt;br&gt;Disagree&lt;br&gt;Neutral&lt;br&gt;Agree&lt;br&gt;Strongly agree</td>
</tr>
<tr>
<td>I felt the Preschool Language Scales-5th Edition was the right choice for evaluating spoken language outcomes for older children in the IHP.</td>
<td>Not applicable&lt;br&gt;Strongly disagree&lt;br&gt;Disagree&lt;br&gt;Neutral&lt;br&gt;Agree&lt;br&gt;Strongly agree</td>
</tr>
</tbody>
</table>
I do not have concerns about the validity/reliability of the Preschool Language Scales-5th Edition

- Not applicable
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

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 IEP:

- Not applicable
- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Please provide any comments related to the program-level outcome monitoring procedures themselves that affected your implementation of program level outcome monitoring procedures over the past year.

__________________________________________________________

Please add any additional comments, questions, or concerns you have about any of your experiences implementing the new program-level outcome monitoring procedures over the past year.

__________________________________________________________
During the pilot, materials were available for download on the OWL site. These were designed to support implementation of the new program-level outcome monitoring procedures. We are interested to know whether you used these materials during the pilot, whether they were useful, and how they could be improved.

<table>
<thead>
<tr>
<th>Description</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>I downloaded or accessed the handout of slides from the webinar to support my use of the new Program Level Outcome Monitoring procedures.</td>
<td>Yes, No</td>
</tr>
<tr>
<td>The handout of slides from the webinar was useful for supporting my use of the new Program Level Outcome Monitoring Procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>I downloaded or accessed the one-page desk reference for SLPs outlining which assessment tools are used at each age for the new Program Level Outcome Monitoring procedures.</td>
<td>Yes, No</td>
</tr>
<tr>
<td>The one-page desk reference for SLPs outlining which assessment tools are used at each age was useful for supporting my use of the new Program Level Outcome Monitoring Procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>I downloaded and shared the one-page summary for parents outlining when Program Level assessments should be completed.</td>
<td>Yes, No</td>
</tr>
<tr>
<td>The one-page summary for parents outlining when assessments should be completed was useful.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>I downloaded or accessed the document to support SLPs in converting raw scores on the MBCDI to percentile ranks.</td>
<td>Yes, No</td>
</tr>
<tr>
<td>The document to support SLPs in converting raw scores on the MBCDI to percentile ranks was useful for supporting my use of the new Program Level Outcome Monitoring Procedures.</td>
<td>Not applicable, Strongly disagree, Disagree, Neutral, Agree, Strongly agree</td>
</tr>
<tr>
<td>I downloaded or accessed the overview document describing the Preschool Language Scales-5th Edition and the transition from the PLS-4 to the PLS-5.</td>
<td>Yes, No</td>
</tr>
<tr>
<td>Question</td>
<td>Not applicable</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>The overview document describing the Preschool Language Scales-5th Edition and the transition from the PLS-4 to the PLS-5 was useful.</td>
<td>○</td>
</tr>
<tr>
<td>I downloaded or accessed the tutorial on the PLS-5 growth scale values.</td>
<td>○ Yes</td>
</tr>
<tr>
<td>The tutorial on the PLS-5 growth scale values was useful for supporting my interpretation of PLS-5 scores.</td>
<td>○ Not applicable</td>
</tr>
<tr>
<td>I downloaded and shared the handout for parents that gave instructions for completing the MCDI.</td>
<td>○ Yes</td>
</tr>
<tr>
<td>The handout for parents that gave instructions for completing the MCDI was useful.</td>
<td>○ Not applicable</td>
</tr>
<tr>
<td>The online learning module was a good way for me to learn how to complete the new Program Level Outcome Monitoring procedures.</td>
<td>○ Not applicable</td>
</tr>
</tbody>
</table>

Please provide any comments related to the implementation materials provided on the OWL site here.

________________________________________
Many SLPs gave suggestions for further supporting implementation of Program Level Outcome Monitoring in the IHP on the surveys completed at the start of the pilot. Please indicate how useful you think each suggestion would be for further supporting implementation of Program Level Outcome Monitoring.

<table>
<thead>
<tr>
<th>Suggestion</th>
<th>Extremely useless</th>
<th>Useless</th>
<th>Neutral</th>
<th>Useful</th>
<th>Extremely useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Include a Frequently Asked Questions Q &amp; A resource to provide direction for commonly identified issues (e.g., what to do when children are English Language Learners ESL and those who are very low functioning).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide additional evidence to demonstrate that the PLS-5 is a valid and reliable tool for assessing preschool children who are deaf or hard of hearing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide additional administrative support or additional time release from clinical duties to facilitate the completion and submission of Program Level outcome monitoring data.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Update documentation for parents to include additional persuasive messaging about the importance of program-level outcome monitoring for monitoring their child's development.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide documentation to clarify that standardized testing is used for reporting program-level outcomes, but SLPs should continue to include multiple forms of assessment (e.g., informal observation, parent interview) as part of their practice.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please provide any comments related to the above suggestions from the pilot survey here. Additional suggestions for improvements to the implementation materials may also be listed.
During the pilot, we collected data for 234 program-level assessments. This included 189 unique children who accessed IHP services, and 43 repeat assessments. With these data, we have been able to demonstrate that the IHP is meeting its goal of ensuring children who are deaf or hard of hearing have spoken language outcomes that are on par with children in other established Early Hearing Detection and Intervention programs, thus supporting the effectiveness of the Ontario IHP. With additional data, we will also be able to show the ways in which children’s spoken language develops during their time in the program.

We also found that standard scores, percentile ranks, and growth scale values from the PLS-5 were accurately converted. This was not the case for data from the MCDI. On the MBCDI, 50% of pilot SLPs made at least one error in their conversion of raw score to percentile rank, and 30% of all percentile calculations were in error. Unfortunately, with this much error, it will be difficult to reliably monitor spoken language outcomes for children in the IHP who are under 19 months of age.

We are requesting your input on how to address this issue (i.e., errors in converting raw scores to percentile rank and percentile calculations) moving forward. We have developed some proposed solutions. Please rate your agreement with the following suggestions.

<table>
<thead>
<tr>
<th>Increase SLPs' awareness of this issue by adding information about the high error rate and types of errors made to the implementation materials.</th>
<th>□ Strongly disagree □ Disagree □ Neutral □ Agree □ Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Include additional training materials to improve SLP accuracy in converting raw scores to percentile ranks on the MBCDI.</td>
<td>□ Strongly disagree □ Disagree □ Neutral □ Agree □ Strongly agree</td>
</tr>
<tr>
<td>Develop a quality assurance test involving MBCDI raw score to percentile rank conversions that SLPs must pass prior to implementation.</td>
<td>□ Strongly disagree □ Disagree □ Neutral □ Agree □ Strongly agree</td>
</tr>
<tr>
<td>Have SLPs tally and report raw scores only for the MBCDI.</td>
<td>□ Strongly disagree □ Disagree □ Neutral □ Agree □ Strongly agree</td>
</tr>
<tr>
<td>Use the PLS-5 as the only tool for monitoring spoken language outcomes in the IHP (i.e., from birth to 6 years of age).</td>
<td>□ Strongly disagree □ Disagree □ Neutral □ Agree □ Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>Increase SLPs' awareness of this issue by adding information about the high error rate and types of errors made to the implementation materials</td>
</tr>
<tr>
<td>2</td>
<td>Include additional training materials to improve SLP accuracy in converting raw scores to percentile ranks on the MBCDI</td>
</tr>
<tr>
<td>3</td>
<td>Develop a quality assurance test involving MBCDI raw score to percentile rank conversions that SLPs must pass prior to implementation</td>
</tr>
<tr>
<td>4</td>
<td>Have SLPs tally and report raw scores only for the MBCDI</td>
</tr>
<tr>
<td>5</td>
<td>Use the PLS-5 as the only tool for monitoring spoken language outcomes in the IHP (i.e., from birth to 5 years of age)</td>
</tr>
</tbody>
</table>

Please provide any comments about the suggestions for addressing the above-noted error rates for the MCDI. If you have additional suggestions for addressing the challenges with accurate reporting of MBCDI scores, please enter them here.
## Appendix 8: SLP’s opinions on the Program-level Outcome Monitoring Procedure

### Statement

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree N (%)</th>
<th>Disagree N (%)</th>
<th>Neither agree nor disagree N (%)</th>
<th>Agree N (%)</th>
<th>Strongly agree N (%)</th>
<th>Mode (Range)</th>
<th>Not applicable N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The IHPs new Program-level Outcome Monitoring Procedure was useful for my clinical practice</td>
<td>1 (2.5%)</td>
<td>4 (10%)</td>
<td>17 (42.5%)</td>
<td>15 (37.5%)</td>
<td>3 (7.5%)</td>
<td>3 (1-5)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>I was able to consistently implement the new Program-level Outcome Monitoring recommendations in my practice.</td>
<td>1 (2.5%)</td>
<td>14 (35%)</td>
<td>7 (17.5%)</td>
<td>17 (42.5%)</td>
<td>1 (2.5%)</td>
<td>4 (1-5)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

## Appendix 9: SLPs opinions of their capacity to implement the Program-level Outcome Monitoring Procedure

### Statement

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

Repeat administration of the Program-level Outcome Monitoring tools to the same child was useful for my own clinical practice.

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

### Appendix 11: SLPs’ Opinions on the quality of the Program-level Outcome Monitoring Procedure

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
<th>Mode (Range)</th>
<th>Not answered</th>
</tr>
</thead>
<tbody>
<tr>
<td>The new Program-level Outcome Monitoring Procedures were similar to the previous outcome monitoring procedures for the IHP.</td>
<td>1 (2.5%)</td>
<td>5 (12.5%)</td>
<td>10 (25%)</td>
<td>23 (57.5%)</td>
<td>2 (2.5%)</td>
<td>4 (1-5)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>The new Program-level Outcome Monitoring Procedures were an improvement over the current procedure.</td>
<td>2 (5%)</td>
<td>4 (5%)</td>
<td>17 (42.5%)</td>
<td>13 (32.5%)</td>
<td>5 (12.5%)</td>
<td>3 (1-5)</td>
<td>1 (2.5%)</td>
</tr>
<tr>
<td>I found the MacArthur-Bates Communicative Development Inventories Words &amp; Gestures to be a high-quality clinical outcome evaluation tool.</td>
<td>1 (2.5%)</td>
<td>6 (15%)</td>
<td>12 (30%)</td>
<td>17 (42.5%)</td>
<td>1 (2.5%)</td>
<td>4 (1-5)</td>
<td>3 (7.5%)</td>
</tr>
<tr>
<td>I found the MacArthur-Bates Communicative Development Inventories Words &amp; Gestures to be a valid and reliable tool for</td>
<td>2 (5%)</td>
<td>4 (10%)</td>
<td>14 (35%)</td>
<td>15 (37.5%)</td>
<td>2 (5%)</td>
<td>4 (1-5)</td>
<td>3 (7.5%)</td>
</tr>
</tbody>
</table>
preschoolers with permanent hearing loss.

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.  

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree N (%)</th>
<th>Disagree N (%)</th>
<th>Neutral N (%)</th>
<th>Agree N (%)</th>
<th>Strongly agree N (%)</th>
<th>Mode (range)</th>
<th>Not applicable N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The IVT procedures were useful for improving services for families of children with hearing loss.</td>
<td>2 (9.5%)</td>
<td>1 (4.7%)</td>
<td>4 (19%)</td>
<td>13 (61%)</td>
<td>0 (0%)</td>
<td>4 (1-4)</td>
<td>3 (12.5%)</td>
</tr>
<tr>
<td>The IVT procedures were useful for my clinical practice.</td>
<td>2 (9.5%)</td>
<td>2 (9.5%)</td>
<td>2 (9.5%)</td>
<td>14 (66.7%)</td>
<td>0 (0%)</td>
<td>4 (1-4)</td>
<td>3 (12.5%)</td>
</tr>
<tr>
<td>I was able to consistently implement the IVT procedures in my practice</td>
<td>3 (13.6%)</td>
<td>5 (22.7%)</td>
<td>3 (13.6%)</td>
<td>10 (45.5%)</td>
<td>0 (0%)</td>
<td>4 (1-4)</td>
<td>2 (8.3%)</td>
</tr>
</tbody>
</table>

Appendix 13: Tier 2 post-pilot survey
IHP End of Pilot - Individual Vulnerability

Thank you for participating in the pilot study to implement new spoken language outcome monitoring procedures in Ontario's Infant Hearing Program (IHP). We appreciate the time and commitment of everyone who supported this initiative and for the valuable data and feedback you have provided thus far.

We are now asking for your feedback after having trialed the new procedures for one year. In this survey, we wish to learn about your experiences with Individual Vulnerability Testing (I.e., administration of 1-2 tests of articulation, vocabulary/grammar, and/or emergent literacy/phonological awareness depending on child age).

Your input is *extremely valued* and will be used to inform decisions about and procedures for province-wide implementation of a new Program-Level Outcome Monitoring process in the IHP. Thank you for sharing your honest perspectives from the front lines.

Note: Please respond to items in this survey considering only your experiences with the Individual Vulnerability Testing sub-pilot study. A separate feedback survey is available about Program Level Outcome Monitoring procedures (i.e., MacArthur Bates Communicative Development Inventories (CDI) and Preschool Language Scales - Fifth Edition (PLS-5)).

It is essential that we be able to link your responses on this survey with the three surveys you completed at the beginning of the pilot. To ensure that you remain anonymous on all of these surveys, we asked you to create an anonymous username at the beginning of the pilot to use for all of the surveys.

As a reminder, this username was required to be at least 8 characters, but was otherwise allowed to contain any number type of characters. We asked you to record your anonymous username somewhere safe and easy for you to access so that you would be able to enter it on future surveys.

Please type that same anonymous username in the box:

Roughly how many IHP children did you attempt to assess using the Individual Vulnerability Testing procedures during the pilot?

- None
- 1 to 5
- 6 to 10
- more than 10

To begin, please answer these three general questions about your experiences with the new Individual Vulnerability Testing.

The Individual Vulnerability Testing procedures were useful for improving services for families of children with hearing loss.

- Not Applicable
- Disagree Strongly
- Disagree
- Neutral
- Agree
- Agree Strongly
<table>
<thead>
<tr>
<th>Question</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Individual Vulnerability Testing procedures were useful for my clinical practice</td>
<td>Not Applicable, Disagree Strongly, Disagree, Neutral, Agree, Agree Strongly</td>
</tr>
<tr>
<td>I was able to consistently implement the Individual Vulnerability Testing procedures in my practice</td>
<td>Not Applicable, Disagree Strongly, Disagree, Neutral, Agree, Agree Strongly</td>
</tr>
</tbody>
</table>

Please select ONE answer for each of the following questions about how your practice environment affected your implementation of Individual Vulnerability Testing procedures over the past year:

<table>
<thead>
<tr>
<th>Question</th>
<th>Response Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>The length of time it took to administer the individual vulnerability tests was appropriate for incorporation into routine clinical practice.</td>
<td>Not Applicable, Disagree Strongly, Disagree, Neutral, Agree, Agree Strongly</td>
</tr>
<tr>
<td>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.</td>
<td>Not Applicable, Disagree Strongly, Disagree, Neutral, Agree, Agree Strongly</td>
</tr>
<tr>
<td>The length of time it took to talk with parents about results of the individual vulnerability tests was appropriate for incorporation into clinical practice.</td>
<td>Not Applicable, Disagree Strongly, Disagree, Neutral, Agree, Agree Strongly</td>
</tr>
<tr>
<td>The time it took to do the individual vulnerability testing and reporting negatively impacted other areas of my practice.</td>
<td>Not Applicable, Disagree Strongly, Disagree, Neutral, Agree, Agree Strongly</td>
</tr>
<tr>
<td>The environment in which I work will make it difficult for me to implement the individual vulnerability testing procedures.</td>
<td>Not Applicable, Disagree Strongly, Disagree, Neutral, Agree, Agree Strongly</td>
</tr>
<tr>
<td>I had the supplies I needed (e.g., test forms) to implement the new individual vulnerability testing procedures.</td>
<td>Not Applicable, Disagree Strongly, Disagree, Neutral, Agree, Agree Strongly</td>
</tr>
<tr>
<td>Statement</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>When I had questions about the Individual Vulnerability Testing procedures, I consulted my colleagues.</td>
<td></td>
</tr>
<tr>
<td>When I had questions about the Individual Vulnerability Testing procedures, I consulted my manager/administrators.</td>
<td></td>
</tr>
<tr>
<td>When I had questions about the Individual Vulnerability Testing procedures, I consulted the &quot;Pilot Implementation Q&amp;A&quot; section of Western's OWL site</td>
<td></td>
</tr>
<tr>
<td>I had the resources I needed (e.g., administrative support for scheduling, data entry) to do the Individual Vulnerability Testing Procedures.</td>
<td></td>
</tr>
<tr>
<td>I had permission from my manager to take the time I needed to complete Individual Vulnerability Testing Procedures.</td>
<td></td>
</tr>
<tr>
<td>Getting timely feedback from experts (e.g., the research team at Western University) helped me to implement the Individual Vulnerability Testing Procedures.</td>
<td></td>
</tr>
<tr>
<td>The SLPs I work with were excited about the new Individual Vulnerability Testing Procedures.</td>
<td></td>
</tr>
<tr>
<td>Managers/administrators I work with were supportive of Individual Vulnerability Testing procedures.</td>
<td></td>
</tr>
</tbody>
</table>
The parents I worked with were interested in the results of individual vulnerability testing procedures.

- Not Applicable
- Disagree Strongly
- Disagree
- Neutral
- Agree
- Agree Strongly

The task of completing the MacArthur-Bates Communicative Development Inventories - Words & Sentences was not too difficult for parents (respondents) to perform.

- Not Applicable
- Disagree Strongly
- Disagree
- Neutral
- Agree
- Agree Strongly

The task of completing the MacArthur-Bates Communicative Development Inventories - Words & Sentences was not too time consuming for parents (respondents) to perform.

- Not Applicable
- Disagree Strongly
- Disagree
- Neutral
- Agree
- Agree Strongly

The task of completing the CELF-P2 Pre-literacy Rating Scale was not too difficult for parents (respondents) to perform.

- Not Applicable
- Disagree Strongly
- Disagree
- Neutral
- Agree
- Agree Strongly

The task of completing the CELF-P2 Pre-literacy Rating Scale was not too time consuming for parents (respondents) to perform.

- Not Applicable
- Disagree Strongly
- Disagree
- Neutral
- Agree
- Agree Strongly

Please provide any comments related to the above questions regarding how your clinical practice environment impacted the implementation of Individual Vulnerability Testing procedures over the past year.

Please select ONE answer for each of the following questions about how your clinical knowledge, skills, and beliefs affected your implementation of Individual Vulnerability Testing procedures over the past year:

Over the past year I felt I had the clinical skills required to implement the new Individual Vulnerability Testing procedures.

- Not Applicable
- Disagree Strongly
- Disagree
- Neutral
- Agree
- Agree Strongly

I am familiar with the administration of the Goldman Fristoe Test of Articulation (GFTA-3).

- Not Applicable
- Disagree Strongly
- Disagree
- Neutral
- Agree
- Agree Strongly
<table>
<thead>
<tr>
<th>I was able to accurately score and use the norms tables for the Goldman Fristoe Test of Articulation (GFTA-3).</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am familiar with the administration of the MacArthur-Bates Communicative Development Inventories - Words &amp; Sentences</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>I was able to accurately score and use the norms tables for the MacArthur-Bates Communicative Development Inventories - Words &amp; Sentences</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>I am familiar with the administration of the Expressive One Word Picture Vocabulary Test - 4th Edition (EOWPVT-4).</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>I was able to accurately score and use the norms tables for the Expressive One Word Picture Vocabulary Test - 4th Edition (EOWPVT-4).</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>I am familiar with the administration of the Clinical Evaluation of Language Fundamentals, Pre-school - Second Edition (CELF-P2) Word Structure subtest.</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>I was able to accurately score and use the norms tables for the Clinical Evaluation of Language Fundamentals, Pre-school - Second Edition (CELF-P2) Word Structure subtest.</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>I am familiar with the administration of the Comprehensive Assessment of Spoken Language - Second Edition (CASL-2) Grammatical Morphemes subtest.</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>I was able to accurately score and use the norms for the Comprehensive Assessment of Spoken Language - Second Edition (CASE-2) Grammatical Morphemes subtest.</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>The new Individual Vulnerability Testing procedures helped with my clinical decision-making.</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>The new Individual Vulnerability Testing procedures helped parents with their decision-making.</td>
<td>Not Applicable</td>
</tr>
<tr>
<td>Repeat administration of the Individual Vulnerability tests to the same child 6-12 months later benefited the families and children that I serve.</td>
<td>Not Applicable</td>
</tr>
</tbody>
</table>

Please provide any comments related to the above questions regarding how your knowledge, skills, or beliefs impacted the implementation of Individual Vulnerability Testing procedures over the past year:

Please select ONE answer for each of the following questions about how the development or content of the individual vulnerability testing procedures impacted how they were implemented over the past year:

<table>
<thead>
<tr>
<th>I found the assessment tools required for Individual Vulnerability Monitoring to be high quality clinical outcome evaluation tools.</th>
<th>Not Applicable</th>
<th>Disagree Strongly</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Agree Strongly</th>
</tr>
</thead>
<tbody>
<tr>
<td>I felt the MacArthur-Bates CDI Words and Gestures &quot;Words Produced&quot; was the right choice for evaluating vocabulary vulnerability in children with permanent hearing loss (6-18 months).</td>
<td>Not Applicable</td>
<td>Disagree Strongly</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Agree Strongly</td>
</tr>
<tr>
<td>I felt the MacArthur-Bates CDI Words and Sentences &quot;Words Produced&quot; was the right choice for evaluating vocabulary vulnerability in children with permanent hearing loss (19-30 months).</td>
<td>Not Applicable</td>
<td>Disagree Strongly</td>
<td>Disagree</td>
<td>Neutral</td>
<td>Agree</td>
<td>Agree Strongly</td>
</tr>
</tbody>
</table>
I felt the Expressive One Word Picture Vocabulary Test (EOWPVT-4) the right choice for evaluating vocabulary vulnerability in children with permanent hearing loss (24-35 months).

☐ Not Applicable
☐ Disagree Strongly
☐ Disagree
☐ Neutral
☐ Agree
☐ Agree Strongly

I felt the CELF-P2 Word Structure subtest the right choice for evaluating grammar vulnerability in children with permanent hearing loss (3-6 years).

☐ Not Applicable
☐ Disagree Strongly
☐ Disagree
☐ Neutral
☐ Agree
☐ Agree Strongly

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).

☐ Not Applicable
☐ Disagree Strongly
☐ Disagree
☐ Neutral
☐ Agree
☐ Agree Strongly

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 (33-48 months).

☐ Not Applicable
☐ Disagree Strongly
☐ Disagree
☐ Neutral
☐ Agree
☐ Agree Strongly

I felt the CELF-P2 Pre-literacy rating scale was the right choice for evaluating emergent literacyphonological awareness vulnerability in children with permanent hearing loss (4-6 years).

☐ Not Applicable
☐ Disagree Strongly
☐ Disagree
☐ Neutral
☐ Agree
☐ Agree Strongly

I felt the CELF-P2 Phonological Awareness subtest was the right choice for evaluating emergent literacyphonological awareness vulnerability in children with permanent hearing loss (4-6 years).

☐ Not Applicable
☐ Disagree Strongly
☐ Disagree
☐ Neutral
☐ Agree
☐ Agree Strongly

I feel the implementation of Individual Vulnerability Testing helped me to identify impairments in children with permanent hearing loss that were missed through Program Level Outcome Monitoring.

☐ Not Applicable
☐ Disagree Strongly
☐ Disagree
☐ Neutral
☐ Agree
☐ Agree Strongly

Please provide any comments related to the individual vulnerability testing procedures themselves that affected your ability to follow the procedures over the past year.

____________________________________________________________

Please add any additional comments, questions, or concerns you have about your experiences implementing Individual Vulnerability Testing procedures over the past year.

____________________________________________________________
During the pilot, materials were available for download on the OWL site. These were designed to support implementation of Individual Vulnerability Testing procedures. We are interested to know whether you used these materials, whether they were useful, and how they could be improved.

<table>
<thead>
<tr>
<th>I downloaded or accessed the handout of slides from the webinar to support my use of Individual Vulnerability Testing procedures.</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>The handout of slides from the webinar was useful for supporting my use of Individual Vulnerability Testing procedures.</td>
<td>Not Applicable</td>
<td>Disagree</td>
</tr>
<tr>
<td>Agree</td>
<td>Agree Strongly</td>
<td></td>
</tr>
<tr>
<td>I downloaded or accessed the one-page desk reference for SLPs outlining which assessment tools are used at each age for Individual Vulnerability Testing.</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>The one-page desk reference for SLPs outlining which assessment tools are to be used at each age was useful for supporting my use of Individual Vulnerability Testing procedures.</td>
<td>Not Applicable</td>
<td>Disagree</td>
</tr>
<tr>
<td>Agree</td>
<td>Agree Strongly</td>
<td></td>
</tr>
<tr>
<td>The online learning module was a good way for me to learn how to complete the new Individual Vulnerability Testing procedures</td>
<td>Not Applicable</td>
<td>Disagree</td>
</tr>
<tr>
<td>Agree</td>
<td>Agree Strongly</td>
<td></td>
</tr>
</tbody>
</table>

Please provide any comments related to the implementation materials provided on the OWL site here.

Many SLPs gave suggestions for further supporting implementation of Individual Vulnerability Testing in the IMP on the surveys completed at the start of the pilot. Please indicate how useful you think each suggestion would be for further supporting implementation of Individual Vulnerability Testing procedures.

<p>| Include a Frequently Asked Questions Q &amp; A resource to provide direction for commonly identified issues (e.g., what to do for children who are English Language Learners ESL and those who are very low functioning). | Extremely useless | Useless | Neutral |
| Useful | Extremely Useful |
| Expand the evaluation of children's articulation skills beyond the age of 4. | Extremely useless | Useless | Neutral |
| Useful | Extremely Useful |</p>
<table>
<thead>
<tr>
<th>Suggestion</th>
<th>Evaluation Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop and disseminate a parent handout explaining the purpose and</td>
<td>Extremely useless, Useless, Neutral, Useful, Extremely Useful</td>
</tr>
<tr>
<td>importance of Individual Vulnerability Testing.</td>
<td></td>
</tr>
<tr>
<td>Provide additional administrative support or additional time release from</td>
<td>Extremely useless, Useless, Neutral, Useful, Extremely Useful</td>
</tr>
<tr>
<td>clinical duties to facilitate the completion of Individual Vulnerability</td>
<td></td>
</tr>
<tr>
<td>Testing.</td>
<td></td>
</tr>
<tr>
<td>Provide additional materials to support clinical use of the tools required</td>
<td>Extremely useless, Useless, Neutral, Useful, Extremely Useful</td>
</tr>
<tr>
<td>for Individual Vulnerability testing (e.g., GFTA-3, CASL-2).</td>
<td></td>
</tr>
<tr>
<td>Provide documentation to clarify that standardized testing is used for</td>
<td>Extremely useless, Useless, Neutral, Useful, Extremely Useful</td>
</tr>
<tr>
<td>monitoring vulnerable areas, but SLPs should continue to include multiple</td>
<td></td>
</tr>
<tr>
<td>forms of assessment (e.g., informal observation, parent interview) as part</td>
<td></td>
</tr>
<tr>
<td>of their practice.</td>
<td></td>
</tr>
<tr>
<td>Eliminate Individual Vulnerability Testing and focus on Program level</td>
<td>Extremely useless, Useless, Neutral, Useful, Extremely Useful</td>
</tr>
<tr>
<td>Outcome Monitoring.</td>
<td></td>
</tr>
</tbody>
</table>

Please provide any comments related to the above suggestions from the subject survey here. Additional suggestions may also be listed.
### Appendix 14: SLPs’ perceptions of time involved in IVT procedure

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree N (%)</th>
<th>Disagree N (%)</th>
<th>Neutral N (%)</th>
<th>Agree N (%)</th>
<th>Strongly agree N (%)</th>
<th>Mode (range)</th>
<th>Not applicable N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The length of time it took to administer the Individual vulnerability tests was appropriate for incorporation into routine clinical practice.</td>
<td>5 (22.7%)</td>
<td>3 (13.6%)</td>
<td>4 (18.2%)</td>
<td>9 (40.9%)</td>
<td>0 (0%)</td>
<td>4 (1-4)</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>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.</td>
<td>2 (9.1%)</td>
<td>4 (18.2%)</td>
<td>3 (13.6%)</td>
<td>12 (54.5%)</td>
<td>0 (0%)</td>
<td>4 (1-4)</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>The length of time it took to talk with parents about results of the Individual vulnerability tests was appropriate for incorporation into clinical practice.</td>
<td>2 (9.5%)</td>
<td>2 (9.5%)</td>
<td>6 (28.6%)</td>
<td>10 (47.6%)</td>
<td>0 (0%)</td>
<td>4 (1-4)</td>
<td>3 (12.5%)</td>
</tr>
<tr>
<td>The time it took to do the Individual vulnerability testing and reporting negatively impacted other areas of my practice.</td>
<td>0 (0%)</td>
<td>8 (36.4%)</td>
<td>8 (36.4%)</td>
<td>4 (18.2%)</td>
<td>1 (4.5%)</td>
<td>3 (2-5)</td>
<td>2 (8.3%)</td>
</tr>
</tbody>
</table>

### Appendix 15: SLPs’ perceptions of practice environment for IVT procedure

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree N (%)</th>
<th>Disagree N (%)</th>
<th>Neutral N (%)</th>
<th>Agree N (%)</th>
<th>Strongly agree N (%)</th>
<th>Mode (range)</th>
<th>Not applicable N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The environment in which I work will made it difficult for me to implement the IVT procedures.</td>
<td>0 (0%)</td>
<td>14 (63.6%)</td>
<td>6 (27.3%)</td>
<td>1 (4.5%)</td>
<td>0 (0%)</td>
<td>2 (2-4)</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>I had the supplies I needed (e.g., test forms) to implement the new IVT procedures.</td>
<td>0 (0%)</td>
<td>1 (4.5%)</td>
<td>0 (0%)</td>
<td>12 (54.5%)</td>
<td>8 (36.4%)</td>
<td>4 (2-5)</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>When I had questions about the IVT procedures, I consulted my colleagues.</td>
<td>0 (0%)</td>
<td>1 (5.3%)</td>
<td>3 (15.8%)</td>
<td>12 (63%)</td>
<td>2 (10.5%)</td>
<td>4 (2-5)</td>
<td>5 (20.8%)</td>
</tr>
<tr>
<td>When I had questions about the IVT procedures, I consulted my manager/administrators.</td>
<td>5 (19%)</td>
<td>1 (5.3%)</td>
<td>7 (36.8%)</td>
<td>3 (15.8%)</td>
<td>7 (36.8%)</td>
<td>2 (1-4)</td>
<td>5 (20.8%)</td>
</tr>
<tr>
<td>When I had questions about the IVT procedures, I consulted the &quot;Pilot Implementation Q&amp;A&quot; section of Western's OWL site.</td>
<td>1 (4.5%)</td>
<td>4 (18.2%)</td>
<td>4 (18.2%)</td>
<td>9 (40.9%)</td>
<td>3 (13.6%)</td>
<td>4 (1-5)</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>I had the resources I needed (e.g., administrative support for scheduling, data entry) to do the IVT Procedures.</td>
<td>3 (21%)</td>
<td>1 (4.8%)</td>
<td>6 (28.6%)</td>
<td>7 (33.3%)</td>
<td>1 (4.8%)</td>
<td>4 (1-5)</td>
<td>3 (12.5%)</td>
</tr>
<tr>
<td>I had permission from my manager to take the time I needed to complete IVT Procedures.</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>5 (23.8%)</td>
<td>13 (61.9%)</td>
<td>2 (9.5%)</td>
<td>4 (3-5)</td>
<td>3 (12.5%)</td>
</tr>
<tr>
<td>Getting timely feedback from experts (e.g., the research team at Western University) helped me to implement the IVT Procedures.</td>
<td>0 (0%)</td>
<td>1 (5.3%)</td>
<td>10 (52.6%)</td>
<td>6 (31.6%)</td>
<td>1 (5.3%)</td>
<td>3 (2-5)</td>
<td>5 (20.8%)</td>
</tr>
<tr>
<td>The SLPs I work with were excited about the new IVT Procedures.</td>
<td>6 (27.3%)</td>
<td>3 (13.6%)</td>
<td>7 (31.8%)</td>
<td>4 (18.2%)</td>
<td>1 (4.5%)</td>
<td>3 (1-5)</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>Managers/administrators I work with were supportive of IVT procedures.</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>5 (23.8%)</td>
<td>14 (66.7%)</td>
<td>1 (4.7%)</td>
<td>4 (3-5)</td>
<td>3 (8.3%)</td>
</tr>
<tr>
<td>The parents I worked with were interested in the results of IVT procedures.</td>
<td>2 (9.1%)</td>
<td>3 (13.6%)</td>
<td>9 (41%)</td>
<td>6 (27.3%)</td>
<td>1 (4.5%)</td>
<td>3 (1-5)</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>The task of completing the MacArthur-</td>
<td>0 (0%)</td>
<td>4 (21%)</td>
<td>3 (11%)</td>
<td>0 (0%)</td>
<td>4 (2-4)</td>
<td>5 (20.8%)</td>
<td></td>
</tr>
</tbody>
</table>
Bates Communicative Development Inventories - Words & Sentences was not too difficult for parents (respondents) to perform.

The task of completing the MacArthur-Bates Communicative Development Inventories - Words & Sentences was not too time consuming for parents (respondents) to perform.

The task of completing the CELF-P2 Pre-literacy Rating Scale was not too difficult for parents (respondents) to perform.

The task of completing the CELF-P2 Pre-literacy Rating Scale was not too time consuming for parents (respondents) to perform.

---

**Appendix 16: SLPs’ opinions of their capacity to implement the IVT procedure**

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

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree N (%)</th>
<th>Disagree N (%)</th>
<th>Neutral N (%)</th>
<th>Agree N (%)</th>
<th>Strongly agree N (%)</th>
<th>Mode (range)</th>
<th>Not applicable N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I found the assessment tools required for the IVT to be high quality clinical outcome evaluation tools.</td>
<td>0 (0%)</td>
<td>1 (4.5%)</td>
<td>3 (13.6%)</td>
<td>14 (64.6%)</td>
<td>3 (13.6%)</td>
<td>4 (2-5)</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>I felt the MacArthur-Bates CDI Words and Gestures &quot;Words Produced&quot; was the right choice for evaluating vocabulary vulnerability in children with permanent hearing loss (8-18 months).</td>
<td>2 (10%)</td>
<td>2 (10%)</td>
<td>4 (20%)</td>
<td>7 (40%)</td>
<td>3 (15%)</td>
<td>4 (1-5)</td>
<td>4 (16.7%)</td>
</tr>
<tr>
<td>I felt the Expressive One Word Picture Vocabulary Test (EOWPVT-4) the right choice for evaluating vocabulary vulnerability in children with permanent hearing loss (24-35 months).</td>
<td>0 (0%)</td>
<td>1 (6.3%)</td>
<td>6 (37.5%)</td>
<td>1 (6.3%)</td>
<td>4 (2-5)</td>
<td>8 (33.3%)</td>
<td></td>
</tr>
<tr>
<td>I felt the CELF-P2 Word Structure subtest the right choice for evaluating grammar vulnerability in children with permanent hearing loss (3-6 years).</td>
<td>0 (0%)</td>
<td>1 (5.3%)</td>
<td>4 (21%)</td>
<td>10 (52.6%)</td>
<td>3 (15.8%)</td>
<td>4 (2-5)</td>
<td>5 (20.8%)</td>
</tr>
<tr>
<td>I felt the Goldman Fristoe Test of Articulation, Third Edition (GFTA-3) - Sounds in Words subtest was the right choice for evaluating vocabulary and grammar vulnerability in children with permanent hearing loss (3-6 years).</td>
<td>1 (5.6%)</td>
<td>3 (16.7%)</td>
<td>3 (16.7%)</td>
<td>7 (38.9%)</td>
<td>3 (16.7%)</td>
<td>4 (1-5)</td>
<td>6 (25%)</td>
</tr>
</tbody>
</table>

The new IVT procedures helped with my clinical decision-making.

The new IVT procedures helped parents with their decision-making.

Repeat administration of the Individual Vulnerability tests to the same child 6-12 months later benefited the families and children that I serve.

**Appendix 17: SLPs’ perceptions of the quality of the IVT procedure**
syntax vulnerability in children with permanent hearing loss (30-48 months). 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).

<table>
<thead>
<tr>
<th>0 (0%)</th>
<th>1 (7.7%)</th>
<th>6 (46%)</th>
<th>5</th>
<th>0 (0%)</th>
<th>3 (2-4)</th>
<th>11 (45.8%)</th>
</tr>
</thead>
</table>

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).

<table>
<thead>
<tr>
<th>0 (0%)</th>
<th>0 (0%)</th>
<th>6 (37.5%)</th>
<th>8 (50%)</th>
<th>1 (6.25%)</th>
<th>4 (3-5)</th>
<th>8 (33%)</th>
</tr>
</thead>
</table>

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.

<table>
<thead>
<tr>
<th>4 (20%)</th>
<th>2 (10%)</th>
<th>1 (5%)</th>
<th>10 (50%)</th>
<th>2 (10%)</th>
<th>4 (1-5)</th>
<th>4 (16.7%)</th>
</tr>
</thead>
</table>

**Appendix 18: Distribution of PLS-5 standard scores**
Appendix 19: The Revised Ottawa Model of Research Use

Description: This figure outlines the Revised Ottawa Model of Research Use

This figure is licensed under a Creative Commons Attribution 4.0 License (https://cjnr.archive.mcgill.ca/article/view/1888/1882). No changes have been made to the original image.
Appendix 20: Vocal development barriers and assessment purposes survey

Confidential

Survey: Vocal Assessment in children with permanent hearing loss - for Speech-Language Pathologists

Researchers at Western University are starting a series of projects to identify tools that assess vocal development and first words to recommend to the Infant Hearing Program (IHP) for use with children who have permanent hearing loss and are younger than 22 months.

In order to develop effective recommendations, it is important to understand your clinical practice context, and identify test purposes that would be useful to your clinical decision making.

This anonymous survey is designed to understand what information you would want to know about the vocal and first word development of a child who has permanent hearing loss and is younger than 22 months if you had the assessment tools capable of answering these questions.

For the purposes of these surveys, we are defining...

1. vocal development as a child's speech sound productions including:
   - Early vocalizations (e.g., cooing, gooking, marginal babble)
   - Age of canonical babble onset
   - Sound repertoire
   - Syllable shapes/complexity

2. first words development as the age at which a child produces their first words and their vocabulary size.

The following questions are designed to understand more about your clinical practice.

How many years have you been practicing as a speech-language pathologist (SLP)?

Do you provide auditory verbal therapy to children with permanent hearing loss?  ○ Yes  ○ No

How many years have you practiced as an SLP in the Preschool Speech & Language (PSL) program?

Do you provide services to children with permanent hearing loss?  ○ Yes  ○ No

Do you provide services to children with permanent hearing loss who are younger than 22 months?  ○ Yes  ○ No

Approximately what proportion of your caseload is children with permanent hearing loss who are younger than 22 months?  ○ None  ○ 25% or less  ○ between 26% and 50%  ○ between 51% and 75%  ○ greater than 76%
<table>
<thead>
<tr>
<th>Question</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>How often do you assess the vocal development of any child (with or without permanent hearing loss) on your caseload when they are younger than 22 months?</td>
<td>Never, Seldom, Sometimes, Frequently, Always</td>
</tr>
<tr>
<td>How often do you assess the first words of any child (with or without permanent hearing loss) on your caseload when they are younger than 22 months?</td>
<td>Never, Seldom, Sometimes, Frequently, Always</td>
</tr>
<tr>
<td>How often do you assess the vocal development of children with permanent hearing loss who are younger than 22 months?</td>
<td>Never, Seldom, Sometimes, Frequently, Always</td>
</tr>
<tr>
<td>How often do you assess the first words of children with permanent hearing loss who are younger than 22 months?</td>
<td>Never, Seldom, Sometimes, Frequently, Always</td>
</tr>
<tr>
<td>Typically, how frequently (e.g., weekly, monthly) do you see a child with permanent hearing loss when they are younger than 22 months?</td>
<td></td>
</tr>
<tr>
<td>It is within my scope of practice as a SLP to assess the vocal development of children who have permanent hearing loss who are younger than 22 months</td>
<td>Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree</td>
</tr>
<tr>
<td>It is within my scope of practice as a SLP to assess the first words of children who have permanent hearing loss who are younger than 22 months</td>
<td>Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree</td>
</tr>
<tr>
<td>I believe that other SLPs regularly assess the vocal development of children who have permanent hearing loss who are younger than 22 months</td>
<td>Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree</td>
</tr>
<tr>
<td>I believe that other SLPs regularly assess the first words of children who have permanent hearing loss who are younger than 22 months</td>
<td>Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree</td>
</tr>
<tr>
<td>In my opinion, families of children with permanent hearing loss who are younger than 22 months want to know about their child’s vocal development</td>
<td>Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree</td>
</tr>
</tbody>
</table>
In my opinion, families of children with permanent hearing loss who are younger than 22 months want to know about their child’s first words development.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

Assessing the vocal development of children with hearing loss who are younger than 22 months provides me with important information.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

Assessing the vocal development of children with permanent hearing loss who are younger than 22 months provides families with important information.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

Assessing the vocal development of children with permanent hearing loss who are younger than 22 months provides audiologists with important information.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

In my current practice, I have the assessment tools I need to conduct an appropriate assessment of a child with permanent hearing loss’ vocal development before 22 months.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

In my current practice, I have the assessment tools I need to conduct an appropriate assessment of a child with permanent hearing loss’ first words before 22 months.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

In my current practice, I have the time I need to conduct an appropriate assessment of a child with permanent hearing loss’ vocal development before 22 months.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

In my current practice, I have the time I need to conduct an appropriate assessment of a child with permanent hearing loss’ first words before 22 months.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

In my current practice, I have the resources (e.g., access to test forms) I need to appropriately assess the vocal development of an child with permanent hearing loss who is younger than 22 months.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

In my current practice, I have the resources (e.g., access to test forms) I need to appropriately assess the first words of a child with permanent hearing loss who is younger than 22 months.

- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree
<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neither agree nor disagree</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have the knowledge I need to conduct an appropriate vocal development assessment of a child with permanent hearing loss who is younger than 22 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I have the knowledge I need to conduct an appropriate first words assessment of a child with permanent hearing loss who is younger than 22 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I have the knowledge I need to appropriately interpret the results of a vocal development assessment of a child with permanent hearing loss who is younger than 22 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I have the knowledge I need to appropriately interpret the results of a first words assessment of a child with permanent hearing loss who is younger than 22 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In this section, you will be presented with a series of sentence prompts and a text box to write statements that complete the prompt. For each prompt, please list as many statements as you can think of that complete the prompt. Indicate the start of a new statement with a "/". If you do not wish to provide statements to a particular prompt, please indicate "No Response"

For example: "In the winter I like to..."

/Go skiing
/Stay inside
/Travel somewhere warm

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>In my clinical practice, I use the results of a child with permanent hearing loss' vocal development to ___</td>
<td></td>
</tr>
<tr>
<td>In my opinion, assessing the vocal development of a child with permanent hearing loss who is younger than 22 months is challenging because ___</td>
<td></td>
</tr>
<tr>
<td>In my opinion, things that would help me assess the vocal development of a child with permanent hearing loss who is younger than 22 months are ___</td>
<td></td>
</tr>
<tr>
<td>Things from a (re)assessment that I need to know about the vocal development of a child with permanent hearing loss who is younger than 22 months are ___</td>
<td></td>
</tr>
<tr>
<td>Things from a (re)assessment that a parent wants to know about their child with permanent hearing loss' vocal development are ___</td>
<td></td>
</tr>
<tr>
<td>Things from a (re)assessment that another professional (e.g., early childhood educator, audiologist) wants to know about a child with permanent hearing loss' vocal development are ___</td>
<td></td>
</tr>
</tbody>
</table>
In this section, you will be presented with a series of clinical questions. All questions pertain to children with permanent hearing loss who are under 22 months. Please indicate, on a scale from 1 to 5 (1 = very unimportant, 5 = very important) how important knowing the answers to the following questions are to you.

<table>
<thead>
<tr>
<th>Question</th>
<th>Very unimportant</th>
<th>Somewhat unimportant</th>
<th>Neutral</th>
<th>Somewhat important</th>
<th>Very important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the child's vocal development within age-expectations compared to children their age who are typically developing and have typical hearing?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the child's vocal development within expectations for children with similar levels of hearing loss?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the child's vocal development within expectations for children with similar amplified hearing levels?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the child's vocal development improved, relative to their same-aged peers, since their last visit?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the child acquired new vocal development abilities since their last visit?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the child's vocal development fallen behind their same-aged peers since their last visit?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the child's vocal development plateaued or not changed since their last visit?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the child's level of vocal development indicate the child is having more problems with language learning than expected based on their hearing loss?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the child's level of vocal development indicate that the child is having more problems with speech development than expected based on their hearing loss?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>Options</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the child's level of vocal development indicate that the child needs more speech and language therapy than they are currently receiving?</td>
<td>Very unimportant, Somewhat unimportant, Neutral, Somewhat important, Very important</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What stage of vocal development has the child mastered?</td>
<td>Very unimportant, Somewhat unimportant, Neutral, Somewhat important, Very important</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What stage of vocal development is emerging?</td>
<td>Very unimportant, Somewhat unimportant, Neutral, Somewhat important, Very important</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Which speech sounds would be appropriate goals for the child?</td>
<td>Very unimportant, Somewhat unimportant, Neutral, Somewhat important, Very important</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Which syllable shapes would be appropriate goals for the child?</td>
<td>Very unimportant, Somewhat unimportant, Neutral, Somewhat important, Very important</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Which words would be appropriate goals the child?</td>
<td>Very unimportant, Somewhat unimportant, Neutral, Somewhat important, Very important</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This section contains clinical questions you might have about a child with permanent hearing loss who is under 22 months. Please pick the 5 that are most important for you to answer. For those not in your top 5, please select “Not in my top 5.”

<table>
<thead>
<tr>
<th>Question</th>
<th>In my top 5</th>
<th>Not in my top 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the child’s vocal development within age expectations compared to children their age who are typically developing and have typical hearing?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the child’s vocal development within expectations for children with similar levels of hearing loss?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is the child’s vocal development within expectations for children with similar amplified hearing levels?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the child’s vocal development improved, relative to their same-aged peers, since their last visit?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the child acquired new vocal development abilities since their last visit?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the child’s vocal development fallen behind their same-aged peers since their last visit?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Has the child’s vocal development plateaued or not changed since their last visit?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the child’s level of vocal development indicate that the child is having more problems with language learning than expected based on their hearing loss?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Does the child’s level of vocal development indicate that the child is having more problems with speech development than expected based on their hearing loss?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>Does the child's level of vocal development indicate that the child needs more speech and language therapy than they are currently receiving?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What stage of vocal development has the child mastered?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What stage of vocal development is emerging?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Which speech sounds would be appropriate goals for the child?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Which syllable shapes would be appropriate goals for the child?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Which words would be appropriate goals for the child?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please note, you have included fewer than 5 clinical questions in your top 5. Please select 5 clinical questions.

Please note, you have included more than 5 clinical questions in your top 5. Please select only 5 clinical questions.

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are there any other important or very important questions that you have about a child's vocal development that were not listed above?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please list the other important or very important questions you have.

Please provide any additional comments.
# Curriculum Vitae

**Name:** Olivia Daub  

**Post-secondary Education and Degrees:**  

<table>
<thead>
<tr>
<th>Year</th>
<th>Degree Type</th>
<th>Institution</th>
<th>Location</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016-2019</td>
<td>MClSc (Speech-Language Pathology) &amp; Ph.D.</td>
<td>The University of Western Ontario</td>
<td>London, Ontario, Canada</td>
<td>2016-2016 MClSc</td>
</tr>
<tr>
<td>2014-2016</td>
<td>MSc</td>
<td>The University of Waterloo</td>
<td>Waterloo, Ontario, Canada</td>
<td>2014-2016 MSc</td>
</tr>
<tr>
<td>2010-2014</td>
<td>BA Psychology</td>
<td>The University of Western Ontario</td>
<td></td>
<td>2014-2014 BA Psychology</td>
</tr>
</tbody>
</table>

**Honours and Awards:**  

<table>
<thead>
<tr>
<th>Year</th>
<th>Award Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-2016</td>
<td>Ontario Graduate Scholarship</td>
</tr>
<tr>
<td>2016-2017</td>
<td>Child Language Research Award</td>
</tr>
<tr>
<td>2018-2019</td>
<td>National Sciences &amp; Engineering Research Council - Undergraduate Student Research Assistantship</td>
</tr>
<tr>
<td>2019-2020</td>
<td>Child Language Research Award</td>
</tr>
<tr>
<td>2020-2021</td>
<td>National Sciences &amp; Engineering Research Council - Undergraduate Student Research Assistantship</td>
</tr>
</tbody>
</table>

**Related Work Experience:**  

<table>
<thead>
<tr>
<th>Role</th>
<th>Course/Experience</th>
<th>Institution</th>
<th>Location</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecturer</td>
<td>COMMSCI 9629S, <em>Clinical Applications in Developmental Speech Sound Disorders</em></td>
<td>The University of Western Ontario</td>
<td></td>
<td>2020, 2021</td>
</tr>
<tr>
<td>Teaching Assistant</td>
<td></td>
<td>The University of Western Ontario</td>
<td></td>
<td>2014-2021</td>
</tr>
</tbody>
</table>

**Peer-Reviewed Publications:**  


White, K.S., & Daub, O. (Submitted). When it’s appropriate not to adapt: Toddlers’ learning of novel speech patterns is affected by visual information. Brain and Language.


Other Scholarly Work:


Daub, O., Cunningham, B.J., Oram Cardy, J. (2020). Individual Spoken Language Vulnerability Monitoring in the Ontario Infant Hearing Program: Results of the


Peer-Reviewed Oral Presentations:


Non Peer-Reviewed Oral Presentations:


presentation to the Ontario MCYS Preschool Speech and Language Program/Infant Hearing Program, Joint Coordinators Meeting, Toronto, Canada.


**Peer-Reviewed Posters:**


