CANADIAN JOURNAL OF SPEECH-LANGUAGE PATHOLOGY & AUDIOLOGY | CJSLPA

Volume 48, No. 1, 2024

REVUE CANADIENNE D'ORTHOPHONIE ET D'AUDIOLOGIE | RCOA

Volume 48, No. 1, 2024

SAC OAC

Speech-Language & Audiology Canada

Orthophonie et Audiologie Canada

Communicating care
La communication à coeur

CJSLPA REVIEWERS

Reviewers for this issue included: Monique Joanne Charest, Maude Desjardins, Jennifer Hatton, Benoît Jutras, Brenda Louw, Eunjung Na, Janis Oram Cardy, Timothy Pommée, and Ann Sutton.

VISION AND MISSION OF SPEECH-LANGUAGE AND AUDIOLOGY CANADA

VISION
The unified national voice of speech-language pathology and audiology.

MISSION
Advancing the professions of speech-language pathology and audiology and empowering members and associates to optimize communication, health and education outcomes for all..

INDEXING

CJSLPA is indexed by:

- CINAHL – Cumulative Index to Nursing and Allied Health Literature
- Elsevier Bibliographic Databases (SCOPUS)
- ProQuest – CSA Linguistics and Language Behavior Abstracts (LLBA)
- PsycInfo
- Thomson Gale – Academic Onefile
- EBSCO Publishing Inc. (CINAHL Plus with full text)
- Directory of Open Access Journals (DOAJ)

ISSN 1913-2020

SCOPE AND PURPOSE OF CJSLPA

SCOPE

The Canadian Journal of Speech-Language Pathology and Audiology (CJSLPA) is a peer-reviewed, online journal of clinical practice for audiologists, speech-language pathologists and researchers.

CJSLPA is an open access journal, which means that all articles are available on the internet to all users immediately upon publication. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose. CJSLPA does not charge publication or processing fees.

PURPOSE

The purpose of CJSLPA is to disseminate current knowledge pertaining to hearing, balance and vestibular function, feeding/swallowing, speech, language and social communication across the lifespan with direct application to Canadian audiology and/or speech-language pathology.

COPYRIGHT

CJSLPA is published by Speech-Language and Audiology Canada (SAC). Publications Agreement Number: # 40036109.
1000-1 Nicholas St., Ottawa, ON K1N 7B7 | 800.259.8519 | www.cjslpa.ca | www.sac-oac.ca

RÉVISEURS ET RÉVISEURES DE LA RCOA

Les personnes suivantes ont agi à titre de réviseur ou réviseure pour ce numéro : Monique Joanne Charest, Maude Desjardins, Jennifer Hatton, Benoît Jutras, Brenda Louw, Eunjung Na, Janis Oram Cardy, Timothy Pommée et Ann Sutton.

VISION ET MISSION D'ORTHOPHONIE ET AUDIOLOGIE CANADA

VISION
Porte-parole de l'orthophonie et de l'audiologie à l'échelle nationale.

MISSION
Faire progresser les professions de l'orthophonie et de l'audiologie et donner aux membres et associés et associées les moyens d'optimiser les résultats en matière de communication, de santé et d'éducation pour tous et toutes.

INDEXATION

La RCOA est indexée dans :

- CINAHL – Cumulative Index to Nursing and Allied Health Literature
- Elsevier Bibliographic Databases (SCOPUS)
- ProQuest – CSA Linguistics and Language Behavior Abstracts (LLBA)
- PsycInfo
- Thomson Gale – Academic Onefile
- EBSCO Publishing Inc. (CINAHL Plus with full text)
- Directory of Open Access Journals (DOAJ)

ISSN 1913-2020

MISSION ET BUT DE LA RCOA

MISSION
La revue canadienne d'orthophonie et d'audiologie (RCOA) est une revue révisée par les pairs sur la pratique clinique, qui est disponible en ligne et qui est destinée aux audiologistes, orthophonistes et chercheurs et chercheuses.

La RCOA est une revue en accès libre, ce qui signifie que tous les articles sont disponibles sur Internet dès leur publication, et ce, pour tous les utilisateurs et toutes les utilisatrices. Les utilisateurs et utilisatrices ont l'autorisation de lire, de télécharger, de copier, de distribuer, d'imprimer, de rechercher ou de fournir le lien vers le contenu intégral des articles, ou encore, d'utiliser les articles à toutes autres fins légales. La RCOA ne charge aucuns frais pour le traitement ou la publication des manuscrits.

BUT
Le but de la RCOA est de diffuser les connaissances actuelles relatives à l'audition, à la fonction vestibulaire et à l'équilibre, à l'alimentation/déglutition, à la parole, au langage et à la communication sociale qui ont une application directe aux domaines de l'orthophonie et de l'audiologie du Canada, et ce, pour tous les âges de la vie.

DROIT D'AUTEUR

La RCOA est publiée par Orthophonie et Audiologie Canada (OAC). Numéro de publication : # 40036109.
1, rue Nicholas, bureau 1000, Ottawa (Ontario) K1N 7B7 | 800.259.8519 | www.cjslpa.ca | www.oac-sac.ca

TABLE OF CONTENTS

TABLE DES MATIÈRES

Évaluation auditivo-perceptive de la résonance et de la parole en lien avec la fonction vélopharyngée : Validation de contenu d'une liste de phrases en français québécois

Auditory-Perceptual Assessment of Resonance and Speech Related to Velopharyngeal Function: Content Validation of a List of Sentences in Québec French

MOTS-CLÉS

RÉSONANCE

PAROLE

FONCTION VÉLOPHARYNGÉE

ÉVALUATION AUDITIVO-PERCEPTIVE

VALIDATION

Marie-Ève Caty
Marianne Paul
Annie Salois
Ericka Beaudoin
Johanie Bouchard
Miroslava Dimova
Lisa Massaro
Andréanne Mayrand
Élisa-Maude McConnell
Eugénie Préfontaine
Alla Sorokin

Marie-Ève Caty [1], Marianne Paul [1], Annie Salois [2], Ericka Beaudoin [3], Johanie Bouchard [3], Miroslava Dimova [2], Lisa Massaro [4], Andréanne Mayrand [2], Élisa-Maude McConnell [3], Eugénie Préfontaine [4] et Alla Sorokin [2]

[1] Département d'orthophonie, Université du Québec à Trois-Rivières, QC, CANADA.

[2] Département d'orthophonie, Centre hospitalier universitaire Sainte-Justine, QC, CANADA.

[3] Département d'orthophonie, Centre hospitalier universitaire de Québec – Université Laval, QC, CANADA.

[4] Département d'orthophonie, Hôpital de Montréal pour enfants – Centre universitaire de santé McGill, QC, CANADA.

Rédacteur :
Vincent Martel-Sauvageau

Abrégé

Le besoin d'un échantillon de parole de qualité pour l'évaluation des individus présentant une dysfonction vélopharyngée est reconnu depuis longtemps. Or, à ce jour, aucune liste de phrases contrôlées n'a été publiée pour l'évaluation auditivo-perceptive de la fonction vélopharyngée en français québécois. Le but du présent projet clinique était donc de développer et de valider une liste de phrases contrôlées en français québécois. Suivant un processus délibératif, ce projet de recherche-action effectué en collaboration avec des orthophonistes expertes dans le domaine des troubles de la résonance et des dysfonctions vélopharyngées s'est effectué en deux phases : (a) la création de la liste de phrases et (b) la validation de contenu des phrases et la documentation de leurs caractéristiques d'utilisation. Lors de la phase 1, les membres d'une communauté de pratique ont développé 20 phrases en suivant des critères phonético-linguistiques spécifiques. La phase 2 a débuté par l'envoi d'un questionnaire pour évaluer le respect et l'exhaustivité des critères pour chacune des phrases et pour documenter les contextes d'utilisation de l'outil. Le questionnaire a été rempli par 13 orthophonistes utilisant la liste de phrases dans leur pratique. Ensuite, deux rencontres de groupe ont été effectuées pour discuter des modifications finales à apporter à la liste. La liste de phrases est considérée par les orthophonistes expertes comme présentant une bonne validité de contenu. Les expertes recommandent l'utilisation de la liste de phrases finale dans le contexte de l'évaluation auditivo-perceptive à partir de trois ou quatre ans et jusqu'à l'âge adulte.

Abstract

It has long been recognized that a quality speech sample is needed to assess individuals with velopharyngeal dysfunction. However, no list of controlled sentences in Québec French has been published for the auditory-perceptual assessment of velopharyngeal function to date. The goal of this clinical project was therefore to develop and validate a list of controlled sentences in Québec French. Using a deliberative inquiry, this action research project was carried out in collaboration with speech-language pathologists with expertise in resonance disorders and velopharyngeal dysfunctions in two phases: (a) creation of the list of sentences and (b) content validation of the sentences and documentation of their usage characteristics. In Phase 1, members of a community of practice developed 20 sentences based on specific phonetic-linguistic criteria. Phase 2 began by sending out a questionnaire to evaluate respect for and completeness of the criteria for each sentence and to document the contexts in which the tool was used. The questionnaire was completed by 13 speech-language pathologists who had used the list of sentences in their practice. Two group meetings were then held to discuss the final modifications to be made to the list. The expert speech-language pathologists considered the sentence list to have good content validity. They also recommend using the final version of the list of sentences when performing auditory-perceptual assessments of individuals aged 3-4 years to adulthood.

Les orthophonistes sont des professionnels qui ont notamment un rôle à jouer dans la prise en charge des individus qui présentent des troubles de la parole en lien avec une dysfonction vélopharyngée ou autres troubles de la résonance. Bien que ce rôle soit prépondérant dans le cadre plus traditionnel des équipes traitant les fentes palatines ou dans les cliniques spécialisées dans l'évaluation de la fonction vélopharyngée, les orthophonistes à l'extérieur de ces équipes peuvent aussi intervenir éventuellement auprès de cette clientèle.

Dysfonction vélopharyngée

La dysfonction vélopharyngée (DVP) est le résultat d'un défaut de l'anatomie et/ou de la physiologie au regard des mouvements du mécanisme vélopharyngé (Kuehn et Moller, 2000; Kummer, 2011b, 2020; Peterson-Falzone et al., 2017). Elle est fréquente chez les enfants nés avec une fente palatine, mais elle peut également être observée chez des enfants qui présentent une anomalie crânio-faciale, un trouble des sons de la parole d'origine neuromotrice (p. ex. dyspraxie, dysarthrie), une hypotonie ou encore un mauvais apprentissage de la fonction vélopharyngée de façon isolée (Peterson-Falzone et al., 2017). La DVP peut entraîner des changements importants quant aux caractéristiques perceptives de la parole sur les plans de la résonance, de l'articulation et/ou de la voix (Kummer, 2011a, 2020). Au sujet de la résonance, l'hypernasalité (voix nasillarde) est une caractéristique de la DVP. En revanche, l'hyponasalité et la résonance cul-de-sac sont généralement reliées à une obstruction des voies respiratoires (p. ex. hypertrophie des végétations adénoïdes ou des amygdales, ou obstruction nasale), et non à une DVP. La DVP peut aussi affecter l'articulation par différentes manifestations liées aux consonnes, telles que de l'émission nasale audible à l'oreille (p. ex. turbulente ou non) ou visible (p. ex. buée sur un miroir nasal), de la faiblesse d'émission causée par une pression intraorale insuffisante sur les consonnes orales ainsi que des mécanismes articulatoires compensatoires (p. ex. coups de glotte ou fricatives pharyngées) (Kummer, 2011a, 2020; Peterson-Falzone et al., 2017). Finalement, la DVP peut entraîner des symptômes vocaux comme une raucité vocale, une faible intensité vocale et/ou une voix éteinte (Cavalli, 2011; Fujiki et Thibeault, 2022). Certains enfants adoptent des comportements vocaux, comme une faible intensité vocale, une voix éteinte ou une fréquence anormalement aiguë, ou encore produisent une grimace nasale (constriction anormale des muscles du visage lors de la production de la parole) pour compenser l'impossibilité de réaliser la fermeture vélopharyngée (Cavalli, 2011; Zajac et Linville, 1989).

Évaluation clinique de la dysfonction vélopharyngée

La qualité de la parole est une mesure importante quant à l'efficacité de l'intervention en orthophonie pour la clientèle présentant une DVP. En clinique, bien qu'elle puisse se réaliser à l'aide d'instruments comme le nasomètre, l'évaluation de la fonction vélopharyngée s'effectue principalement par les orthophonistes via l'évaluation auditivo-perceptive, soit « à l'oreille » (Baylis et al., 2019; Henningsson et al., 2008; Kuehn et Moller, 2000; Kummer, 2011a; Sell, 2005). Cette évaluation est d'ailleurs reconnue comme étant la norme (le *gold standard*) pour juger de la qualité de la parole (Sell, 2005). Durant cette évaluation, l'orthophoniste analyse toutes les caractéristiques de la parole énumérées précédemment à l'aide d'échantillons de parole. Ceux-ci peuvent être obtenus au moyen de différentes tâches : répétition de syllabes, de mots ou de phrases, productions de séries automatiques (p. ex. jours de la semaine ou comptage), conversation ou lecture de textes à voix haute (Baylis et al., 2019; Cummings et Bae, 2018).

La répétition de phrases contrôlées demeure une tâche de choix pour les orthophonistes afin de s'assurer de la validité et de la fiabilité de l'évaluation auditivo-perceptive, en particulier auprès des non-lecteurs. En effet, cette tâche permet d'évaluer rapidement l'articulation, l'émission nasale et la résonance dans un contexte de parole continue plus représentatif de la parole spontanée que la répétition de mots isolés ou de syllabes. Pour contrôler les phrases, il importe de vérifier la fréquence des voyelles fermées (aussi dites hautes) dans la phrase. En effet, il est reconnu que ces dernières présentent davantage d'hypernasalité que les voyelles ouvertes (aussi dites basses), car leur position linguale est plus élevée et qu'une plus grande force est requise lors de la fermeture vélopharyngée (Kuehn et Moon, 1998; Lewis et al., 2000; Moon et al., 1994). Pour ce qui est des consonnes, il est nécessaire de contrôler la présence de consonnes à forte pression intraorale, c'est-à-dire les consonnes qui nécessitent une accumulation d'air avant leur relâchement soudain (occlusives) ou graduel (fricatives), puisque ces dernières peuvent se retrouver plus affectées en présence d'une DVP (Karnell et al., 2001; Kummer, 2020). Ainsi, tant en clinique qu'en recherche, pour une évaluation valide et fiable, l'utilisation d'un ensemble de phrases contrôlées dans une tâche de répétition est à privilégier.

Des listes de phrases contrôlées en français européen sont publiées (Garnier, 2012; Todic et al., 2022; Tourmel, 2012). Elles ont été développées pour l'établissement de normes de nasalance pour l'évaluation instrumentale avec le nasomètre. Ces listes de phrases ne ciblent donc pas

toutes les consonnes du français, dans le but de garder le protocole d'évaluation à l'aide du nasomètre le plus court possible pour faciliter son utilisation en clinique. Par contre, lors de l'évaluation auditivo-perceptive, il importe d'utiliser des phrases ciblant toutes les consonnes de la langue d'intérêt, car certaines caractéristiques de la parole ou de la résonance peuvent être spécifiques à un phonème/une consonne. Rappelons que le français québécois contient 36 phonèmes, soit 17 consonnes (14 orales et 3 nasales), 3 semi-consonnes et 16 voyelles (12 orales et 4 nasales; Martin, 1996). De plus, le français québécois présente des traits phonétiques spécifiques tels que l'affrication, la réduction de groupes consonantiques de consonnes finales, le relâchement vocalique en syllabe fermée en fin de mot et la diphtongaison, pour ne nommer que ceux-ci (Ostiguy et Tousignant, 2008; Reinke et Ostiguy, 2016). Or, une liste de phrases qui sont contrôlées en français québécois et qui ciblent toutes les consonnes de cette langue n'existe pas à ce jour.

Lignes directrices pour développer des phrases contrôlées pour l'évaluation auditivo-perceptive de la résonance et de la parole en lien avec la fonction vélopharyngée

Henningsson et al. (2008; Hutters et Henningsson, 2004) ont élaboré des lignes directrices pour développer des phrases contrôlées afin d'obtenir un échantillon de parole valide et fiable. Pour évaluer l'hypernasalité, les émissions nasales et/ou turbulences nasales et les erreurs de production des consonnes, il est suggéré d'échantillonner les consonnes voisées et non voisées à forte pression intraorale de la langue parlée dans laquelle on souhaite créer les stimuli. Pour évaluer l'hyponasalité, ils recommandent que les phrases soient chargées d'un mélange de consonnes nasales. Ces lignes directrices pour construire des stimuli contrôlés permettent non seulement d'obtenir des échantillons de parole adéquats pour l'évaluation auditivo-perceptive, mais aussi de rapporter les résultats à l'aide de paramètres universels qui permettent la comparaison des résultats obtenus entre établissements et aussi entre différentes langues (Henningsson et al., 2008).

À partir de ces lignes directrices, Trost-Cardamone (2012) a développé une série de 26 phrases contrôlées en anglais nommées *American English Sentence Sample*, et par la suite, a émis des recommandations supplémentaires pour la création de phrases contrôlées. Pour les phrases évaluant l'hypernasalité et les erreurs de production des consonnes, Trost-Cardamone a proposé d'ajouter des phrases avec des consonnes à faible pression intraorale. De plus, pour évaluer l'hyponasalité, elle a suggéré d'ajouter, pour chaque consonne nasale de l'inventaire phonémique de la langue considérée, une phrase chargée uniquement d'une de ces consonnes et de n'avoir qu'une seule phrase incluant toutes les consonnes nasales. Finalement, elle a ajouté une phrase qui contient un mélange de consonnes nasales et de consonnes à forte pression et à faible pression pour évaluer l'influence des consonnes nasales sur les consonnes orales. Le **tableau 1** présente les critères phonético-linguistiques de Trost-Cardamone (2012), inspirés de ceux d'Henningsson et al. (2008), opérationnalisés pour l'inventaire phonémique du français québécois.

Comme mentionné précédemment, à ce jour, aucune liste de phrases contrôlées n'a été publiée pour l'évaluation auditivo-perceptive de la fonction vélopharyngée en français québécois. C'est pourquoi une communauté de pratique regroupant des orthophonistes travaillant auprès de cette clientèle a entrepris d'en développer une.

Objectif de recherche

L'objectif du projet de recherche est donc de créer une liste de phrases contrôlées en français québécois, d'en évaluer la validité de contenu et de documenter ses caractéristiques d'utilisation. Ces caractéristiques concernent les aspects perceptifs de la parole pouvant être évalués par les différentes catégories de phrases de cette liste de même que les conditions liées à son administration. Un objectif secondaire était d'évaluer l'exhaustivité des critères phonético-linguistiques de l'anglais pour leur application à l'évaluation en français québécois. Pour ce faire, nous nous sommes inspirées de l'approche pour le développement d'outils d'évaluation proposée par Boateng et al. (2018) étant donné la pertinence de son organisation générale pour le présent projet. Deux phases ont donc été effectuées : (a) la création des phrases et leur prétest afin d'identifier lesquelles conserver et (b) la validation du contenu des phrases ainsi que la documentation de leurs caractéristiques d'utilisation.

Ce projet de recherche est aussi inspiré du courant de la recherche-action qui se veut une orientation pragmatique, démocratique et participative à la création de connaissances réunissant l'action et la réflexion de même que la théorie et la pratique (Bradbury, 2015). Ce type de recherche reconnaît que les experts du terrain, notamment les professionnels de la santé, possèdent un savoir expérientiel et des compétences utiles afin d'identifier des problèmes dans la pratique, de mettre en œuvre des solutions, de surveiller le processus de changement et d'en évaluer les résultats sur le plan de la transformation des pratiques (Clavreul et Albuquerque, 2020). Ainsi, l'expertise clinique a été valorisée tout au long des phases de création et de validation de contenu des phrases contrôlées en français québécois.

Tableau 1

Critères phonético-linguistiques utilisés pour l'élaboration des phrases selon leur catégorie et leur opérationnalisation en français québécois (adaptés de Henningsson et al. [2008] et Trost-Cardamone [2012])

Catégories de phrase	Critères	Opérationnalisation en français québécois
Consonnes		
1. Phrases avec des consonnes orales à forte pression intra-orale	**1.1** Chaque phrase doit être « chargée »[a] seulement d'une consonne cible orale à forte pression.	Les consonnes orales sont : p, t, k, b, d, g, f, s, ʃ, v, z, ʒ, ʁ
	1.2 La consonne ciblée se retrouve dans toutes les positions d'occurrence de la langue.	Les trois positions considérées sont : initiale, médiane et finale.
	1.3 Les consonnes non-ciblées doivent être : a) à faible pression ou b) à forte pression, mais avoir le même lieu d'articulation que la consonne ciblée.	Les consonnes orales à faible pression sont : l, w, j, ɥ[b] Les consonnes orales à forte pression par lieu d'articulation : • Labiales : p, b, f, v • Coronales : t, d, s, z, ʃ, ʒ • Vélaires : k, g • Uvulaire : ʁ
	1.4 Aucune phrase n'a de consonne nasale.	Les consonnes nasales sont : m, n, ɲ
2. Phrases avec des consonnes orales à faible pression intra-orale	**2.1** Chaque phrase inclut seulement des consonnes orales à faible pression.	Les consonnes orales à faible pression sont : l, w, j, ɥ[b]
3. Phrases avec des consonnes nasales	**3.1** Chaque phrase doit être « chargée »[a] uniquement d'une des consonnes nasales.	Les consonnes nasales sont : m, n, ɲ
	3.2 Une phrase doit contenir un mélange des consonnes nasales.	Les consonnes nasales sont : m, n, ɲ[c]
	3.3 Les consonnes non-ciblées doivent être à faible pression.	Les consonnes orales à faible pression sont : l, w, j, ɥ[b]
4. Phrase avec un mélange des trois types de consonnes	**4.1** La phrase inclut les trois types de consonnes, soit celles orales à forte pression, orales à faible pression et nasales.	Les consonnes orales à forte pression sont : p, t, k, b, d, g, f, s, ʃ, v, z, ʒ, ʁ Les consonnes orales à faible pression sont : l, w, j, ɥ[b] Les consonnes nasales sont : m, n, ɲ
Voyelles		
Pour chacune des 4 catégories de phrases précédentes	La phrase inclut une variété de voyelles orales, notamment sur le plan du degré d'ouverture. Pour les catégories 3 et 4, les voyelles nasales sont aussi permises[d].	Types de voyelles orales (et nasales) par degré d'ouverture : • Fermées - non arrondie : i - arrondies : y, u • Mi-fermées à mi-ouvertes : e, ø, o, ə, ɛ, œ, ɔ (ẽ/ɛ[e], œ̃, ɔ̃) • Ouvertes, arrondies ou non : a, ɑ, (ã/ɑ̃[e])

[a] La phrase est considérée « chargée » si elle contient une très haute concentration de la consonne ciblée qui peut donc être identifiée facilement. [b] Dans ce contexte, /w, j, ɥ/ sont considérées comme des consonnes, bien que dans les faits elles sont plutôt des semi-consonnes/semi-voyelles. [c] Étant donné sa faible fréquence d'occurrence en français québécois, ce phonème n'a pas été utilisé dans la création de la phrase. [d] En anglais, il n'y a pas de voyelle nasale. Ceci est donc une adaptation du critère pour le français québécois qui possède quatre voyelles nasales. [e] Les deux variantes sont possibles pour cette voyelle nasale en français québécois (Martin et al., 2001).

Pour les deux phases du projet, une certification éthique a été demandée, mais ne s'est pas révélée requise pour procéder à ce projet de recherche. En effet, il fut jugé que les participantes-orthophonistes n'étaient pas personnellement visées par la recherche, mais plutôt sollicitées pour leur opinion d'expertes.

Méthode de la phase 1 : Création de la liste de phrases

Participants

Les expertes consultées étaient des orthophonistes spécialisées dans l'évaluation des individus présentant une dysfonction vélopharyngée provenant de trois centres hospitaliers universitaires québécois. Ces orthophonistes étaient toutes membres d'une communauté de pratique en dysfonction vélopharyngée et malformations orofaciales mise sur pied en 2012. Les membres de cette communauté sont considérés comme des expertes dans leur domaine. La composition de la communauté de pratique change au fil des années, notamment en raison de membres qui changent de champ d'exercice. La phase 1 fut menée par 11 orthophonistes, dont 8 ont été présentes tout au long de cette phase.

Procédure

La création de la liste de phrases fut menée entre 2013 et 2014. Il a été décidé d'utiliser les critères phonético-linguistiques suggérés par Henningsson et al. (2008) et par Trost-Cardamone (2012). Conformément à ces critères, quatre catégories de phrases variant selon les caractéristiques des consonnes de la langue ciblée ont été créées. Ces catégories, les critères pour chacune de celles-ci et leur opérationnalisation pour tenir compte des spécificités de l'inventaire phonémique en français québécois sont présentés au **tableau 1**. Par exemple, le son « r » comporte plusieurs variantes et celle choisie a été la fricative uvulaire /ʁ/ qui est la plus fréquente en français québécois (Martin, 1996; Reinke et Ostiguy, 2016) alors qu'en anglais la variante utilisée est considérée comme liquide (Trost-Cardamone, 2012). En se basant sur ces critères, toutes les membres de la communauté de pratique ont développé des phrases sémantiquement signifiantes à l'aide d'un processus de co-construction visant le consensus et laissant place à la créativité et à l'expérience de chacune. À l'occasion, avant de procéder à l'acceptation des phrases finales, un travail complémentaire d'analyse ou de révision en sous-groupe fut réalisé. Lors de la création des phrases, l'implantation de l'étape de réduction d'items proposée par Boateng et al. (2018) fut modifiée, car jugée peu efficace dans le contexte. En effet, puisque chaque phrase doit correspondre à des critères spécifiques, il aurait fallu créer plus d'une version de chaque phrase pour ensuite sélectionner la meilleure. Il a plutôt été décidé de modifier les phrases jugées plus faibles à la suite de l'utilisation des phrases en clinique par les expertes, pour les améliorer. Ainsi, une première liste de phrases a été créée (voir les phrases préliminaires présentées dans le **tableau 2**). Une première utilisation clinique de cette liste a été réalisée par les orthophonistes de chaque centre dans le cadre de leur pratique régulière d'évaluation auprès de jeunes de moins de 18 ans, occasionnellement jusqu'à 21 ans, présentant une dysfonction vélopharyngée. À la lumière des observations cliniques des orthophonistes, des ajustements ont été apportés à quatre phrases afin de les rendre plus significatives sémantiquement ou moins difficiles à prononcer d'un point de vue articulatoire (voir les phrases ajustées présentées dans le **tableau 2**).

Finalement, il fut décidé d'illustrer les phrases ajustées afin d'avoir un support visuel à présenter aux jeunes enfants pour faciliter la répétition. Ces illustrations sont en couleur, relativement épurées et présentent des personnages de type bande dessinée.

Résultats de la phase 1

Le résultat de ce travail a permis d'obtenir une liste de 20 phrases contrôlées en français québécois pouvant être utilisées lors de l'évaluation auditivo-perceptive de la résonance et de la parole en lien avec la dysfonction vélopharyngée (voir les phrases ajustées présentées dans le **tableau 2**). Ces phrases ont alors été rendues disponibles aux membres de la communauté de pratique pour une utilisation dans leur pratique.

Méthode de la phase 2 : Validation du contenu des phrases et documentation de leurs caractéristiques d'utilisation

Après quelques années d'utilisation, l'intérêt grandissant pour ces phrases a poussé la communauté de pratique à poursuivre le travail de développement de la liste de phrases en entamant un travail de validation de contenu. Pour ce faire, la deuxième auteure (M. Paul), une chercheuse se spécialisant dans le développement et la validation d'outils, a été sollicitée. La deuxième phase a donc consisté à valider le contenu des phrases développées à la phase 1, à obtenir de l'information sur leurs caractéristiques d'utilisation et à évaluer l'exhaustivité des critères phonético-linguistiques pour le français québécois. Pour ce faire, un processus délibératif a été utilisé avec un volet quantitatif suivi d'un volet qualitatif. Le processus délibératif est une approche utilisée en recherche-action et se situe à mi-chemin entre la méthode Delphi et le groupe de discussion (Kanuka 2010; Wouters et al., 2019). En effet, à l'instar de la méthode Delphi, les participants doivent délibérer sur une ou plusieurs questions visant à explorer

Tableau 2

Évolution de la formulation des phrases

Cibles	Phrases préliminaires version 2013	Phrases ajustées version 2014[a]	Phrases finales version 2020
Phrases de la catégorie 1			
1. /p/	Papa paie la loupe.	Papa paie la loupe.	Papa paie la loupe.
2. /b/	Le bel habit à Bob.	Le bel habit à Bob.	Le bel habit à Bob.
3. /t/	Le toutou a la tête haute.	Le toutou a la tête haute.	Le toutou a la tête haute.
4. /d/	Dalie aide l'ado.	**La doudou est laide.**	La doudou est laide.
5. /k/	Coco le coq est coquet.	Coco le coq est coquet.	Coco le coq **a le kiwi.**
6. /g/	Hugues a le LEGO® et la glue.	**Guy** a le LEGO® **à Hugues**.	Guy a le LEGO® à Hugues.
7. /f/	Fifi la fée a l'œuf.	Fifi la fée a l'œuf.	Fifi la fée a l'œuf.
8. /v/	Ève lave le veau.	Ève lave le veau.	**Vivi** lave le veau.
9. /s/	Le loup a sucé la suce.	**Ce** loup a sucé la suce.	Ce loup a sucé la suce.
10. /z/	Les oiseaux lisent au zoo.	Les oiseaux lisent au zoo.	Les oiseaux lisent au zoo.
11. /ʃ/	Le chat lèche le hochet.	Le chat lèche le hochet.	**Chili** le chat lèche le hochet.
12. /ʒ/	Gigi a la luge.	Gigi a la luge.	Gigi a la luge.
13. /ʁ/	Ariel et Laure rient.	**Alors**, Arielle **a** ri.	Alors, Arielle a ri.
Phrases de la catégorie 2			
14. /l/	Lili a lu à l'île.	Lili a lu à l'île.	**Éli** a lu à l'île.
15. /w, j/	Wow! Louis a le yoyo.	Wow! Louis a le yoyo.	Wow! Louis a le yoyo.
Phrases de la catégorie 3			
16. /n/	Nina a la lune.	Nina a la lune.	Nina a la lune.
17. /m/	Maman aime le miel.	Maman aime le miel.	Maman aime le miel.
18. /ɲ/	L'agneau aime l'oignon.	L'agneau aime l'oignon.	L'agneau **éloigne** l'oignon.
19. /n, m/	Mon ami a un long nez.	Mon ami a un long nez.	Mon ami a un long nez.
Phrases de la catégorie 4			
20. variées	Blanche-Neige pique son bâton brun.	Blanche-Neige pique son bâton brun.	Blanche-Neige pique son bâton brun.
S1. variées	-	-	**Il a planté des sapins**[b].
S2. variées	-	-	**Elle aime bien le chocolat**[b].

Note. **En gras** : les mots qui ont été changés ou les phrases ajoutées lors des différentes périodes de révision des phrases. S1 et S2 = phrases supplémentaires.
[a] Abel et al. (2015, 2016). [b] Phrase tirée du protocole d'évaluation de la nasalance de Garnier (2012) et de Tourmel (2012).

les points de divergence d'opinions afin de parvenir à une entente qui guidera la prise de décision. Il s'apparente également à un groupe de discussion, puisqu'il implique la collecte de données qualitatives lors d'entretien de groupe. En favorisant la concertation entre les différentes parties prenantes, le processus délibératif permet la co-interprétation des résultats, la co-construction de solutions issues du croisement des savoirs de chacun et ultimement, la prise de décision plus éclairée (Gauvin, 2009). Ce processus repose sur le présupposé que les décisions sont socialement construites et naissent de la discussion (Kelland et Kanuka 2007; Wouters et al., 2019).

Participants

Volet quantitatif

Toutes les orthophonistes (*N* = 13) membres de la communauté de pratique au printemps 2020 ont reçu un courriel les invitant à remplir un questionnaire en ligne à propos de la validation du contenu et de la documentation des caractéristiques de la liste de phrases. Certaines de ces orthophonistes avaient participé à la phase 1 en totalité (*n* = 5) ou en partie (*n* = 3), alors que d'autres n'y avaient pas participé (*n* = 5). Ces orthophonistes étaient toutes des femmes, l'âge moyen était de 36 ans (*É.-T.* = 6,7; *n* = 12) et elles présentaient des profils variés quant à leur ancienneté et à leur expérience d'utilisation de la liste de phrases (voir le **tableau 3**).

Volet qualitatif

À la suite du volet quantitatif, les membres de la communauté de pratique (celles ayant participé au volet quantitatif et une nouvelle membre) ont été invitées à deux rencontres de groupe. Un total de 11 orthophonistes étaient présentes à la première rencontre et 8 orthophonistes étaient présentes à la deuxième rencontre qui a eu lieu un mois plus tard. Les deux chercheuses M.-È. Caty et M. Paul ont coanimé ces deux rencontres d'une durée de trois heures chacune.

Collecte de données

Volet quantitatif

Un questionnaire comprenant cinq sections a été développé et était composé principalement de questions évaluées avec une échelle de type Likert à cinq items (p. ex. « Ne respecte pas du tout le critère » à « Respecte parfaitement bien le critère »). Il y avait aussi comme sixième choix la possibilité de ne pas répondre (p. ex. choisir la réponse « Je ne sais pas »). Le questionnaire comportait

Tableau 3

Caractéristiques socio-démographiques des orthophonistes (*N* = 13) ayant participé au questionnaire en ligne (volet quantitatif)

	n	Proportion (%)
Sexe		
Femme	13	100,0
Ancienneté comme orthophoniste		
0-1 an	1	7,7
2-5 ans	2	15,4
6-10 ans	5	38,5
11-15 ans	4	30,8
16-20 ans	0	0,0
Plus de 20 ans	1	7,7
Ancienneté dans le domaine des DVP et autres troubles de la résonance		
0-1 an	4	30,8
2-5 ans	2	15,4
6-10 ans	6	46,2
11-15 ans	1	7,7
16-20 ans	0	0,0
Plus de 20 ans	0	0,0
Expérience dans l'utilisation de l'outil		
Moins de 3 mois	1	7,7
4-24 mois	4	30,8
3-5 ans	4	30,8
Plus de 5 ans	4	30,8

Note. DVP = dysfonction vélopharyngée

aussi des questions ouvertes d'approfondissement facultatives. Les répondantes avaient ainsi la possibilité de justifier leurs choix et de consigner leurs commentaires et suggestions à la fin de chaque section. La première section portait sur les informations sociodémographiques des répondantes (cinq questions). La deuxième section portait sur la validation de contenu, soit le respect et l'exhaustivité des critères d'élaboration des phrases de l'outil. Pour évaluer le respect des critères, il y avait une question par critère (13 questions, divisées selon les 4 catégories de phrases). Par exemple, pour les 13 phrases de la catégorie 1, les répondantes devaient juger si chaque phrase respectait les 5 critères pour cette catégorie. Pour l'évaluation de l'exhaustivité des critères, la question suivante fut posée aux répondantes pour chacune des catégories (quatre questions) : « Pour la ou les phrase(s) de cette catégorie, pensez-vous que d'autres critères auraient dû être considérés? » La troisième section comportait quatre questions visant à documenter l'utilité de chacune des catégories de phrases pour l'évaluation de la résonance et de la parole en lien avec la fonction vélopharyngée. Par exemple, les répondantes devaient déterminer leur degré d'accord avec la proposition suivante : « Les 2 phrases de la catégorie 2 (phrases contenant uniquement des consonnes orales nécessitant une faible pression) permettent de bien évaluer l'hypernasalité ». La quatrième section comprenait huit questions au sujet des caractéristiques liées à son administration. Par exemple, les répondantes devaient donner leur opinion au sujet de l'ordre de passation et de la durée en fonction de l'âge des patients, de la fréquence d'utilisation des phrases dans différents contextes (p. ex. l'évaluation auditivo-perceptive, l'évaluation avec le nasomètre ou l'évaluation en vidéofluoroscopie ou en nasoendoscopie). La cinquième et dernière section comportait une question facultative qui visait à recueillir des commentaires sur le questionnaire et des commentaires généraux à propos de l'outil. Au total, le questionnaire comportait 34 questions obligatoires et 27 questions facultatives.

Volet qualitatif

Un guide pour les entretiens de groupe sous forme d'un PowerPoint a été développé. On y présentait les résultats du volet quantitatif et des questions permettant d'approfondir ou de justifier ces résultats. L'utilisation des résultats anonymes du questionnaire a permis aux orthophonistes de partager librement leurs points de vue et d'évaluer les contributions des autres membres de manière impartiale, sans préjugés basés sur l'auteure des commentaires. Les co-animatrices ont géré les points divergents dans les résultats du volet quantitatif avec neutralité en nommant les désaccords, en invitant les participantes ayant des points de vue différents à prendre la parole et en utilisant la relation respectueuse et positive entre les orthophonistes pour faire avancer les discussions. Ceci a permis d'éviter le phénomène de la pensée de groupe (*groupthink*; DiPierro et al., 2022), c'est-à-dire de la pression liée à la conformité dans un groupe. Ainsi, ce processus a mis à profit l'expérience collective des orthophonistes afin d'arriver à des décisions consensuelles soutenues par l'autorité de l'ensemble du groupe.

Procédures et analyse des données

Volet quantitatif

Le sondage a été diffusé du 26 mai 2020 au 13 juillet 2020 via la plateforme *Lime Survey*. Un rappel a été effectué un mois après l'invitation initiale.

Les données ont été extraites et analysées grâce à des statistiques descriptives à l'aide du logiciel Excel (version 2016, pour Macintosh). Pour la validation de contenu, l'analyse s'est basée sur le calcul d'indices de validité de contenu pour chaque énoncé (IVC-É; Désormeaux-Moreau et al, 2020; Lynn, 1986; Rubio et al., 2003). Les IVC-É sont calculés, pour chacune des phrases des différents critères, en divisant le nombre d'expertes qui ont répondu « Respecte bien le critère » ou « Respecte parfaitement bien le critère » par le nombre total d'expertes ayant répondu. L'IVC-É moyen de chaque catégorie de phrases a été déterminé en calculant la moyenne des IVC-É obtenus aux phrases de la catégorie. Il est suggéré qu'un seuil d'acceptabilité de 0,78 soit utilisé pour juger de la représentativité des items (Lynn, 1986; Polit et al., 2007). Une révision des items est effectuée si l'IVC-É est inférieur à ce seuil; parfois, le retrait d'items s'impose si l'IVC-É est très bas. Finalement, tous les résultats quantitatifs obtenus ont été interprétés à la lumière d'une analyse qualitative des commentaires laissés par les expertes. Pour documenter l'utilité de chacune des catégories de phrases pour l'évaluation auditivo-perceptive, un calcul de proportion a été effectué en fonction du nombre de répondantes qui ont répondu « D'accord » ou « Totalement d'accord ». Une absence de réponse était considérée comme un score de zéro. Pour être jugée utile, une catégorie devait obtenir une proportion de 10 à 12 des 12 répondantes, l'utilité était jugée limitée avec 8 à 9 des 12 répondantes et la catégorie était jugée peu utile avec moins de 8 des 12 répondantes.

Volet qualitatif

Ce volet visait à discuter des résultats obtenus dans le volet quantitatif. Les deux rencontres de groupe ont été effectuées par visioconférence via la plateforme Zoom.

Les objectifs de ces rencontres étaient de : (a) discuter et modifier les phrases jugées à corriger et (b) échanger à propos des caractéristiques d'utilisation de la liste de phrases. Lors de la première rencontre, des règles de fonctionnement ont d'abord été établies et il fut convenu que les décisions seraient prises de façon consensuelle. Quand le consensus ne pouvait être établi à la suite de la période de discussion, un vote était alors effectué et l'obtention de la majorité était nécessaire. Les points de discussion et les décisions définitives prises de même que leurs justifications ont été documentés en temps réel. Une analyse interprétative a été utilisée pour explorer les résultats du volet quantitatif à la lumière des résultats du volet qualitatif et intégrer les résultats de ces deux volets (Pluye, 2019).

Résultats de la phase 2

Les résultats intègrent les volets quantitatif et qualitatif.

Validation de contenu : Respect et exhaustivité des critères pour la création des phrases

Les phrases ont été analysées une à une pour vérifier si elles respectent les critères de leur catégorie. Pour chacune des catégories, un nombre variable de critères pour les consonnes devait être respecté (voir le **tableau 1**). Concernant les voyelles, le même critère de diversité était appliqué pour toutes les catégories. Les résultats pour chacune des catégories sont présentés dans les prochains paragraphes, suivis des résultats quant à l'exhaustivité.

Phrases de la catégorie 1 (consonnes à forte pression intraorale)

La catégorie 1 compte 13 phrases qui devaient respecter 4 critères pour les consonnes en plus du critère de variété des voyelles (voir le **tableau 1**). Pour le critère 1.1, l'IVC-É moyen est 0,99 (*É.-T.* = 0,02; min = 0,92; max = 1,00). Une seule répondante a jugé que la phrase 9 ne respectait pas suffisamment le critère, mais aucune justification n'a été fournie. À la suite des rencontres de groupe, il fut expliqué que c'est en raison de l'absence de la consonne cible « s » dans le mot de contenu « loup ». Il avait cependant été décidé par les auteures originales de la liste de limiter l'utilisation de consonnes non ciblées à faible pression aux mots de fonction. Or, puisque le « l » de loup peut être considéré comme une consonne non ciblée à faible pression et que cela respecte le critère 1.3 (voir le **tableau 1**), cette déviation fut jugée mineure par les expertes. La phrase 9 n'a donc pas été modifiée.

Pour le critère 1.2, l'IVC-É moyen est 0,96 (*É.-T.* = 0,54; min = 0,54; max = 1,00). Six répondantes ont jugé que la phrase 8 ne respectait pas le critère. L'analyse qualitative des justifications explique ce résultat par l'absence du phonème /v/ en position médiane. À la suite des rencontres de groupe, la phrase 8 a été modifiée.

Pour les critères 1.3 et 1.4, les IVC-É moyens sont de 1,00 (tous les IVC-É sont de 1,00). En effet, l'ensemble des répondantes a jugé que toutes les phrases répondaient parfaitement à chaque critère. Aucune phrase n'a donc été modifiée au regard de ces critères.

Concernant le critère pour les voyelles, l'IVC-É moyen est 0,85 (*É.-T.* = 0,22; min = 0,31; max = 1,00). Bien que les phrases 1, 10 et 12 présentaient des ICV-É supérieurs à 0,78, elles ont été jugées par une ou deux personnes comme ne répondant pas au critère. Les phrases 5, 8 et 11, quant à elles, présentaient des IVC-É inférieurs à 0,78 et ont été jugées par six ou neuf personnes comme répondant au critère. Les répondantes justifient qu'une variété de voyelles n'est pas présente dans ces phrases. Lors des rencontres de groupe, une discussion sur le concept de variété de voyelles a d'abord été effectuée et a permis de raffiner le critère, suivant la proposition émise par Henningsson et al. (2008) à ce sujet : s'il est préférable d'avoir les trois types de voyelles orales, il est acceptable de n'en avoir que deux types si une voyelle fermée est incluse. Les phrases 10 et 12 n'ont pas été changées puisqu'elles répondaient à ce critère modifié. La phrase 1, qui avait été initialement considérée comme manquant un type de voyelle, n'a pas été changée non plus, car à la suite d'une analyse approfondie, il fut constaté que les trois types de voyelles étaient bien présents. Les phrases 5, 8 et 11 ont été modifiées à la suite des discussions.

Phrases de la catégorie 2 (consonnes à faible pression intraorale)

La catégorie 2 contient deux phrases qui devaient respecter un critère pour les consonnes et le critère pour les voyelles (voir le **tableau 1**). Pour le critère 2.1, l'IVC-É moyen est 1,00 (tous les IVC-É sont de 1,00). En effet, les répondantes ont jugé que toutes les phrases répondaient parfaitement bien à ce critère. Concernant le critère pour les voyelles, l'IVC-É moyen est 0,85 (*É.-T.* = 0,08; min = 0,77; max = 0,92). La phrase 14 présente un résultat un peu plus faible avec un IVC-É de 0,77. L'analyse qualitative des justifications des répondantes indique qu'il manque un type de voyelles pour la phrase 14. À la suite des rencontres de groupe, seule la phrase 14 a été modifiée.

Phrases de la catégorie 3 (consonnes nasales)

La catégorie 3 contient quatre phrases, soit une phrase pour chacune des trois consonnes nasales du français québécois et une incluant un mélange de ces trois consonnes. Ces phrases devaient répondre à trois critères pour les consonnes en plus du critère de variété de

voyelles (voir le **tableau 1**). Pour le critère 3.1, qui concerne les trois premières phrases, l'IVC-É moyen est 0,97 (*É.-T.* = 0,04; min = 0,92; max = 1,00). Seule la phrase 18 a été jugée ne pas respecter suffisamment le critère par une répondante, car la consonne cible n'est pas présente en position finale de mot. À la suite des rencontres de groupe, la phrase 18 a été modifiée.

Pour le critère 3.2, qui concerne la phrase 19 (incluant un mélange des trois consonnes nasales), l'IVC-É est de 0,15. En effet, 11 répondantes ont jugé que le critère n'était pas respecté. L'analyse qualitative des justifications a permis d'expliquer ce résultat par le manque du phonème /ɲ/. Or, lors des rencontres de groupe, cette absence a été jugée non problématique par les expertes. En effet, selon leur expérience, les consonnes « m » et « n » sont les meilleures pour évaluer l'hyponasalité et devraient donc être favorisées pour augmenter la valeur évaluative de cette phrase. L'absence du phonème /ɲ/ a peu d'effet sur la représentativité des consonnes nasales dans la phrase étant donné la très faible fréquence d'occurrence de ce son en français québécois. En effet, dans un corpus à l'oral de 692 707 phonèmes, le /ɲ/ a une fréquence d'occurrence de seulement 721, comparativement à une fréquence de 21 358 pour le /n/ et de 29 476 pour le /m/ (Bédard et al., 2017). Pour sa part, la variante /ŋ/ utilisée lors de la prononciation des mots empruntés de l'anglais, généralement ceux dont la terminaison est « –ing » (Walter, 1983), présente une fréquence d'utilisation de 93 (Bédard et al., 2017). À la suite des rencontres de groupe, la phrase 19 n'a donc pas été modifiée.

Pour le critère 3.3, qui concerne les quatre phrases ciblées dans la catégorie 3, l'IVC-É moyen est de 0,98 (*É.-T.* = 0,03; min = 0,92; max = 1,00). Seule la phrase 18 a été jugée par une répondante comme ne respectant pas suffisamment le critère, car elle inclut la consonne nasale « m » qui n'est pas ciblée. À la suite des rencontres de groupe, la phrase 18 a donc été modifiée.

Finalement, concernant le critère pour les voyelles, l'IVC-É moyen est de 0,88 (*É.-T.* = 0,09; min = 0,77; max = 1,00). La phrase 16, avec un IVC-É de 0,77, ressort comme un peu plus faible, car une voyelle de type mi-fermée à mi-ouverte est manquante. De même, pour les phrases 17 et 18, malgré des IVC-É respectifs de 0,92 et 0,85, il ressort de cela dans les commentaires offerts par les répondantes que tous les types de voyelles ne sont pas représentés dans ces phrases et qu'en général, seulement deux des trois types le sont. Tout comme pour les phrases de la catégorie 1, le critère pour les voyelles a été raffiné lors des rencontres de groupe. En effet, il a été jugé par les expertes que dans ce contexte de phrases nasales, le concept de variété de voyelles porte autant sur le trait de nasalité que sur le degré d'ouverture. Ainsi, il est acceptable d'avoir seulement deux types de voyelles tant qu'une voyelle orale fermée ou une voyelle nasale est incluse. Finalement, aucune phrase n'a été modifiée.

Phrase de la catégorie 4 (mixte, soit avec trois types de consonnes)

La catégorie 4 ne contient qu'une seule phrase qui devait respecter un critère pour les consonnes et le critère pour la variété de voyelles (voir le **tableau 1**). Pour le critère 4.1 et le critère pour les voyelles, les IVC-É sont de 1,00. En effet, les répondantes ont jugé que la phrase 20 respectait bien les deux critères. Donc, la phrase 20 n'a pas été modifiée lors de la rencontre de groupe.

Exhaustivité des critères d'élaboration des phrases

Pour chacune des quatre catégories de phrases, les répondantes devaient se prononcer sur l'exhaustivité des critères phonético-linguistiques, c'est-à-dire indiquer si elles pensaient qu'un ou des critères additionnels auraient dû être considérés. Pour les quatre catégories, aucun critère additionnel n'a été proposé. Cependant, lors des rencontres de groupe, le critère portant sur la variété de voyelles pour les phrases 3 et 4 a été nuancé pour prendre en considération les voyelles nasales. L'opérationnalisation du critère 3.2 concernant le mélange des consonnes nasales (voir le **tableau 1**) a aussi été adaptée pour tenir compte de la faible fréquence d'occurrence du /ɲ/ en français québécois.

En somme, 5 des 20 phrases ont été modifiées (voir les phrases finales 5, 8, 11, 14 et 18 du **tableau 2**), dont 4 pour obtenir une variété parfaite de voyelles. Lors des rencontres de groupe, outre la prise en considération de l'ensemble des critères à respecter, la facilité à répéter la phrase pour un jeune enfant et le fait que l'action représentée dans la phrase soit plausible et puisse facilement être illustrée ont aussi été pris en considération pour procéder à la modification des phrases. Au total, trois des phrases modifiées ont entraîné le changement des illustrations. Il fut convenu de profiter de cette occasion pour augmenter la diversité ethnoculturelle et limiter les stéréotypes des personnages qui y sont représentés.

Caractéristiques d'utilisation : Aspects à considérer lors de l'évaluation auditivo-perceptive selon les catégories de phrases

Les aspects de la résonance et de la parole en lien avec la fonction vélopharyngée à évaluer lors de l'évaluation auditivo-perceptive ont été documentés pour chacune des catégories de phrases. Une des répondantes a été exclue

des analyses de cette section du questionnaire, car elle n'avait pas répondu aux questions en raison de son manque d'expérience avec l'utilisation des phrases. Cette section compte donc 12 répondantes.

Selon les répondantes, les phrases de la catégorie 1 sont utiles pour évaluer : (a) la faiblesse d'émission (manque de pression intraorale), (b) la présence de mécanismes d'articulation compensatoires et (c) la présence d'émissions nasales audibles (turbulentes ou non turbulentes). En effet, toutes les répondantes (12/12) étaient « D'accord » ou « Totalement d'accord » pour utiliser les phrases de la catégorie 1 pour ces objectifs d'évaluation. Pour 83 % des répondantes (10/12 des répondantes), les phrases de cette catégorie s'avèrent aussi utiles pour évaluer l'hypernasalité. Lors des rencontres de groupe, il fut précisé que cette catégorie permet principalement de confirmer la présence de difficultés sévères.

Les phrases de la catégorie 2 ont été jugées utiles par les répondantes pour évaluer : (a) l'hypernasalité (10/12 des répondantes) et (b) la faiblesse d'émission (manque de pression intraorale) (11/12 des répondantes). De plus, pour 75 % des répondantes (9/12 des répondantes), les phrases de la catégorie 2 avaient une utilité limitée pour évaluer la présence de mécanismes d'articulation compensatoires et d'émission nasale. Lors des rencontres de groupe, il a été précisé que l'hypernasalité est généralement plus saillante sur les consonnes voisées à faible pression. Par conséquent, les phrases de cette catégorie s'avèrent particulièrement utiles pour détecter les cas légers d'hypernasalité. Toutefois, il fut reconnu que l'échantillon demeure réduit pour évaluer cette caractéristique puisque la catégorie ne contient que deux phrases. Il fut aussi discuté que les phrases de la catégorie 2 étaient moins utiles que celles de la catégorie 1 pour évaluer la présence de mécanismes d'articulation compensatoires et d'émission nasale. En effet, les mécanismes d'articulation compensatoires sont plus rarement présents sur les consonnes à faible pression et les déperditions nasales peuvent être visibles (à l'aide d'un miroir) pour les consonnes à faible pression, mais elles sont rarement audibles.

Les phrases de la catégorie 3, quant à elles, sont considérées comme utiles par 92 % des répondantes (11/12 des répondantes) pour évaluer : (a) l'hyponasalité et (b) la dénasalisation des consonnes. Les phrases de cette catégorie n'ont pas été jugées utiles pour évaluer l'hypernasalité ni la présence de mécanismes d'articulation compensatoires et d'émission nasale, avec respectivement 2/12, 4/12, et 1/12 des répondantes qui étaient « D'accord » ou « Totalement d'accord ». Lors des rencontres de groupe, il fut précisé que l'hypernasalité est plus difficile à percevoir lorsqu'il y a dominance de phonèmes nasaux, car il y a un effet de contamination jugé normal dans un tel contexte. De plus, il fut souligné que les mécanismes d'articulation compensatoires ne se produisent pas sur les consonnes nasales et que les émissions nasales sont attendues pour les phonèmes nasaux.

Pour la phrase de la catégorie 4, elle a été jugée utile pour évaluer : (a) la faiblesse d'émission (manque de pression intraorale), (b) l'hyponasalité, (c) la présence de mécanismes d'articulation compensatoires et (d) la présence d'émission nasale, avec respectivement 11/12, 10/12, 11/12 et 10/12 des répondantes qui étaient « D'accord » ou « Totalement d'accord ». Lors des rencontres de groupe, il fut conclu que les phrases de la catégorie 1 demeurent plus utiles pour l'évaluation de ces caractéristiques puisque la catégorie 4 ne comporte qu'une seule phrase. De plus, bien que la phrase soit plus représentative de la parole spontanée, le mélange de consonnes orales et nasales complique l'évaluation de la résonance et la détection d'erreurs sur le plan de l'articulation.

Bien qu'il ait été jugé adéquat de ne conserver qu'une seule phrase pour la catégorie 4, à la suite des rencontres de groupe, il a tout de même été décidé d'ajouter deux phrases – identifiées comme supplémentaires (S1 et S2) – tirées des travaux de Garnier (2012) et Tourmel (2012). Ces phrases comportent un pourcentage de phonèmes nasaux inférieur à la phrase mixte déjà incluse, ce qui a été considéré par les expertes comme un ajout intéressant pour augmenter la variété des phrases mixtes. L'évaluation ad hoc de ces deux phrases indique que les critères, dans leur version modifiée, sont bien respectés.

Concernant la présence de particularités sur le plan de la voix, caractéristique discutée lors des rencontres de groupe seulement, aucune des catégories de phrases n'est jugée adéquate pour évaluer cette caractéristique, l'échantillon spontané demeurant la meilleure façon de l'évaluer.

Caractéristiques d'utilisation : Conditions liées à l'administration de la liste de phrases

Finalement, les résultats quantitatifs portant sur les caractéristiques liées à l'administration de la liste de phrases ont été analysés. Ces derniers ont été présentés lors des rencontres de groupe pour les valider, et nous présentons dans cette section les résultats intégrés. La liste de phrases fait consensus quant à son utilité pour l'évaluation auditivo-perceptive, l'évaluation en vidéofluoroscopie, l'évaluation en nasoendoscopie, ainsi que pour documenter l'efficacité de l'intervention (pré et post). Par contre, il n'est pas recommandé d'utiliser ces phrases lors de l'évaluation avec le nasomètre, car il n'y a pas de normes de nasalance pour ces

phrases à ce jour. Il n'est pas recommandé non plus d'utiliser ces phrases en contexte d'intervention, car il ne s'agit pas de phrases fonctionnelles. De plus, les utiliser en contexte d'intervention conduirait à un apprentissage des phrases les rendant ensuite inutilisables pour juger de l'efficacité de l'intervention.

De plus, il fut conclu que ces phrases peuvent être utilisées à partir de l'âge de trois ou quatre ans et sont pertinentes jusqu'à l'âge adulte. Toutefois, à partir de l'adolescence, le contenu des phrases pourrait être jugé trop enfantin. Quant au temps requis pour administrer l'ensemble des phrases, il est estimé à une durée entre 10 et 15 minutes pour les enfants d'âge préscolaire, à environ 10 minutes pour les enfants d'âge scolaire et à une durée entre 5 et 10 minutes pour les adolescents et les adultes. Néanmoins, le temps de passation peut être augmenté de quelques minutes si l'orthophoniste réévalue certaines phrases avec le tube auditif ou le miroir, si un renforcement est utilisé ou si la collaboration ou l'attention de la personne évaluée est faible.

Finalement, la totalité des orthophonistes répondantes (n = 13) mentionne adapter le nombre de phrases utilisées lors de l'évaluation en fonction de l'âge de la personne évaluée, de ses habiletés sur le plan de la parole et du langage, de sa collaboration et du temps imparti. Globalement, l'ordre de passation des phrases est jugé comme peu important, les phrases pouvant être administrées dans l'ordre voulu.

Discussion

L'objectif de la présente étude était de créer une liste de phrases contrôlées en français québécois, d'en évaluer la validité de contenu et de documenter ses caractéristiques d'utilisation. Au total, la liste validée comprend 20 phrases contrôlées en français québécois et deux phrases supplémentaires qui ont fait l'objet d'un consensus. Les aspects perceptifs pouvant être évalués par les différentes catégories de phrases sont conformes à ceux retrouvés dans la littérature au sujet de l'évaluation des caractéristiques perceptives de résonance et de la parole liée à la fonction vélopharyngée (Kummer, 2011a; Lewis et al., 2000; Marino et al., 2020). Quant aux caractéristiques liées à l'administration de la liste de phrases, ces dernières se veulent des informations importantes qui se retrouveront dans le manuel de passation qui accompagnera les phrases.

Cette liste de phrases contrôlées favorise une évaluation auditivo-perceptive valide et fiable en permettant l'évaluation de l'articulation, des émissions nasales et de la résonance dans un contexte plus représentatif de la parole spontanée que la répétition de mots isolés ou de syllabes. Le contrôle des phrases a permis d'ajuster le type et la fréquence des voyelles et des consonnes incluses pour assurer une bonne représentation des caractéristiques perceptives du français québécois avec un nombre limité de phrases. En conversation spontanée, avant d'obtenir un échantillon de parole permettant d'évaluer tous les sons, il faut généralement collecter et analyser un échantillon beaucoup plus long.

Les résultats obtenus mettent en lumière la nécessité d'adapter certains des critères développés pour l'anglais, dans le but de tenir compte des caractéristiques de l'inventaire phonémique du français québécois. Trois adaptations ont été effectuées dans le présent projet. La première touche le /r/ anglophone, une consonne liquide considérée comme à faible pression dans la liste de phrases de Trost-Cardamone (2012), qui a été remplacé par le /ʁ/, une fricative uvulaire considérée comme une consonne à forte pression (Martin, 1996; Reinke et Ostiguy, 2016) et plus utilisée en français québécois. La deuxième adaptation concerne le critère de variété de voyelles. Ce dernier fut nuancé pour que les voyelles nasales soient incluses, car le français québécois en possède quatre, contrairement à l'anglais qui n'en possède pas. Le critère de variété de voyelle a aussi été ajusté pour permettre l'inclusion de seulement deux des trois types de voyelles orales, si une voyelle fermée est incluse. La troisième adaptation porte sur le phonème /ɲ/ qui n'a pas été retenu dans la construction de la phrase incluant un mélange des consonnes nasales en raison de sa très faible fréquence d'occurence en français québécois. Ces trois adaptations soulignent l'importance de procéder à l'ajustement des critères phonético-linguistiques et de réfléchir à leur opérationnalisation, comme dans le cas de la présente étude, pour s'assurer d'inclure les caractéristiques pertinentes de la langue dans laquelle l'outil sera adapté.

L'évaluation auditivo-perceptive demeure la norme de référence (le *gold standard*; Kuehn et Moller, 2000) pour établir le diagnostic de trouble de la parole en lien avec une dysfonction vélopharyngée ou tout autre trouble de la résonance, mais elle demeure un défi pour tous et toutes les orthophonistes. En effet, plusieurs facteurs peuvent influencer la fiabilité de l'évaluation auditivo-perceptive dont le type de stimuli utilisé, la présence d'un trouble de la voix ou de l'articulation cooccurrent, la pertinence des échelles de notation, et la formation et l'expérience de l'orthophoniste de même que ses caractéristiques personnelles telles que son profil linguistique et son âge (Lee et al., 2020). Concernant le type de stimuli utilisé, la conversation spontanée a longtemps été considérée comme étant le type le plus valide, mais sa variabilité

intrinsèque nuit à la fiabilité interjuge dans l'évaluation. En revanche, la tâche de répétition de phrases permet de contrôler le contexte phonétique des stimuli, ce qui facilite l'analyse perceptive et favorise l'accord intra- et interjuge (Klintö et al., 2011; Marino et al., 2020; Sell, 2005). Il est d'ailleurs recommandé que chaque langue ait une liste de phrases contrôlées et approuvées par l'objet d'un consensus (Sell, 2005), et c'est l'absence d'une telle liste en français québécois qui a motivé le présent projet.

Les 20 phrases contrôlées phonétiquement développées et validées dans la présente étude permettront d'utiliser des échelles de notation développées pour l'évaluation auditivo-perceptive des enfants nés avec une fente palatine telles que les échelles de notation liées aux paramètres universels d'Henningsson et al. (2008). Bien que les échelles de notation soient publiées en anglais, elles peuvent être utilisées pour l'évaluation en français québécois puisque, comme son nom l'indique, elles sont liées à des paramètres universels caractérisant la parole, peu importe la ou les langues parlées. Leur utilisation est recommandée pour toutes les langues, tant qu'elles sont utilisées avec des phrases contrôlées pour la ou les langues ciblées (Henningsson et al, 2008). Il est aussi envisageable d'utiliser les phrases avec les échelles de notation du *The cleft audit protocol for speech—augmented* développées par John et al. (2006). Ainsi, les phrases contrôlées en français québécois pourront être utilisées en combinaison avec des échelles de notation en clinique ou en recherche.

Limites de l'étude et pistes de recherche futures

La principale limite de l'étude réside dans l'inclusion d'orthophonistes qui avaient contribué à la création initiale de la liste de phrases dans le groupe, ce qui a pu entraîner un biais favorable envers les phrases. Cette limite est en partie compensée par la présence d'autres orthophonistes qui n'y avaient pas contribué. Aussi, les usagers n'ont pas été sondés dans le cadre du projet, mais la consultation des expertes a permis de tenir compte de façon indirecte de leurs réactions. Que les phrases finales n'aient pas été réévaluées de manière formelle constitue une autre limite de l'étude, bien que tous les critères aient été pris en considération lors des modifications apportées.

De futures études pourront porter sur des critères non documentés liés à la validité, à la fidélité et à des données normatives, ce qui nécessite l'utilisation de l'outil auprès de participants. Il demeure néanmoins approprié de recommander l'utilisation de la liste de phrases finale, puisqu'aucune alternative actuellement disponible ne remplit, à notre connaissance, le critère de validité de contenu. Il serait aussi pertinent de procéder à la traduction, à l'adaptation et à la validation transculturelle (Corbière et Fraccaroli, 2020) des échelles de notation.

Conclusion

Ce projet a permis de concevoir et de procéder à la validation de contenu d'une liste de 20 phrases qui tiennent compte des spécificités du français québécois. Ces phrases ont été développées par un groupe d'expertes selon des critères phonético-linguistiques adaptés à partir de ceux suggérés par Henningsson et al. (2008) et Trost-Cardamone (2012). Elles peuvent être utilisées dès maintenant par les orthophonistes pour l'évaluation auditivo-perceptive de la résonance et de la parole en lien avec la fonction vélopharyngée auprès d'enfants dès l'âge de trois ou quatre ans et jusqu'à l'âge adulte. Ce projet a aussi permis de mettre en lumière trois adaptations qui peuvent être requises lors de l'opérationnalisation des critères phonético-linguistiques dans une langue autre que l'anglais. La prochaine étape de validation de la liste est de procéder à l'évaluation de la fidélité interjuge en utilisant un processus de recherche collaborative et participative entre chercheurs et cliniciens, car ce processus, bien que plus long, permet de mieux intégrer les savoirs liés à l'expertise clinique.

Références

Abel, K., Beaudoin, E., Bouchard, J., Caty, M.-È., Dimova, M., Erdos, C., Lacour, S., Massaro, L., Mayrand, A., McConnell, É.-M., Salois, A. et Sorokin, A. (2016, 16-17 septembre). *Standardization of a protocol for the assessment of Québécois French-speaking individuals with velopharyngeal dysfunction - Part 1: Development of a sentence set* [communication par affiche]. 1st International Symposium on Velopharyngeal Dysfunction, Columbus, Ohio, USA.

Abel, K., Beaudoin, E., Bouchard, J., Caty, M.-È., Dimova, M., Erdos, C., Lacour, S., Massaro, L., Mayrand, A., McConnell, É.-M., Salois, A. et Sorokin, A. (2015, 20-25 avril). *Standardization of a protocol for the assessment of Québécois French-speaking individuals with velopharyngeal dysfunction - Part 1: Development of a sentence set* [communication par affiche]. American Cleft Palate-Craniofacial Association 72nd Annual Meeting and Symposium, Palm Springs, CA, USA.

Baylis, A. L., Perry, J., Wilson, K., Dailey, S., Bedwinek, A., Trost-Cardamone, J., Kirschner, R. E. et Goudy, S. (2019). Team management of velopharyngeal inadequacy: Practical suggestions for speech-language pathologists and surgeons. *Perspectives of the ASHA Special Interest Groups, 4*(5), 850-856. https://doi.org/10.1044/2019_PERS-SIG5-2019-0003

Bédard, P., Audet, A.-M., Drouin, P., Roy, J.-P., Rivard, J. et Tremblay, P. (2017). SyllabO+: A new tool to study sublexical phenomena in spoken Quebec French. *Behavior Research Methods, 49*, 1852-1863. https://doi.org/10.3758/s13428-016-0829-7

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R. et Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*(149), 1-18. https://doi.org/10.3389/fpubh.2018.00149

Bradbury, H. (2015). *The SAGE handbook of action research* (3e éd.). SAGE Publications. https://doi.org/10.4135/9781473921290

Cavalli, L. (2011). Voice assessment and intervention. Dans S. Howard et A. Lohmander (dir.), *Cleft palate speech : Assessment and intervention* (p. 181-198). John Wiley & Sons. https://doi.org/10.1002/9781118785065.ch10

Clavreul, H. et Albuquerque, S. (2020). La recherche-action, une démarche méthodologique pour renforcer la pratique. *Revue Francophone de Recherche en Ergothérapie, 6*(1), 93-102. https://doi.org/10.13096/rfre.v6n1.172

Coenen, H. (2001). Recherche-action : rapports entre chercheurs et acteurs. *Revue internationale de psychosociologie*, VII, 19-32. https://doi.org/10.3917/rips.016.0019

Corbière, M. et Fraccaroli, F. (2020). La conception, la validation, la traduction et l'adaptation transculturelle d'outils de mesure : Des exemples en santé mentale et travail. Dans M. Corbière et N. Larivière (dir.) *Méthodes qualitatives, quantitatives et mixtes, 2e édition : Dans la recherche en sciences humaines, sociales et de la santé* (2e éd., p. 703-752). Presses de l'Université du Québec. https://doi.org/10.2307/j.ctv1c29qz7.32

Cummings, C. et Bae, Y. (2018). Synthesizing speech and imaging findings during the assessment of velopharyngeal function. *Perspectives of the ASHA Special Interest Groups, 3*(5), 48-63. https://doi.org/10.1044/persp3.SIG5.48

Désormeaux-Moreau, M., Naud-Lepage, G. et Drolet, M.-J. (2020). Assurer la validité de contenu lors du développement d'un outil d'évaluation : Soutenir l'ergothérapeute pas à pas. *Revue Francophone de Recherche en Ergothérapie, 6*(2), 75-97. https://doi.org/10.13096/rfre.v6n2.178

DiPierro, K., Lee, H., Pain, K. J., Durning, S. J. et Choi, J. J. (2022). Groupthink among health professional teams in patient care: A scoping review. *Medical Teacher, 44*(3), 309-318. https://doi.org/10.1080/0142159X.2021.1987404

Fujiki, R. B. et Thibeault, S. L. (2022). Are children with cleft palate at increased risk for laryngeal pathology? *The Cleft Palate Craniofacial Journal, 60*(11), 1385-1394. https://doi.org/10.1177/10556656221104027

Garnier, E. (2012). Établissement de normes de nasalance sur une population adulte de langue maternelle française ne présentant pas de rhinolalie [mémoire de maîtrise, Université de Genève]. Archive ouverte UNIGE. http://archive-ouverte.unige.ch/unige:23489

Gauvin, F.-P. (2009). *Qu'est-ce qu'un processus délibératif? Fiche d'information -Pour des connaissances en matière de politiques publiques favorables à la santé.* Centre de collaboration nationale sur les politiques publiques et la santé. https://www.inspq.qc.ca/sites/default/files/publications/1192_processusdeliberatif.pdf

Henningsson, G., Kuehn, D. P., Sell, D., Sweeney, T., Trost-Cardamone, J. E. et Whitehill, T. L. (2008). Universal parameters for reporting speech outcomes in individuals with cleft palate. *The Cleft Palate Craniofacial Journal, 45*(1), 1-17. https://doi.org/10.1597/06-086.1

Hutters, B. et Henningsson, G. (2004). Speech outcome following treatment in cross-linguistic cleft palate studies: Methodological implications. *The Cleft Palate Craniofacial Journal, 41*(5), 544-549. https://doi.org/10.1597/02-164.1

John, A., Sell, D., Sweeney, T., Harding-Bell, A. et Williams, A. (2006). The cleft audit protocol for speech—augmented: A validated and reliable measure for auditing cleft speech. *The Cleft Palate Craniofacial Journal, 43*(3), 272-288. https://doi.org/10.1597/04-141.1

Kanuka, H. (2010). Deliberative inquiry. Dans M. Savin-Baden et C. H. Major (dir.). *New approaches to qualitative research: Wisdom and uncertainty* (1re éd., p.100-108). Routledge.

Kelland, J. H. et Kanuka, H. (2007). "We just disagree:" Using deliberative inquiry to seek consensus about the effects of e-learning on higher education. *Canadian Journal of Learning and Technology, 33*(3), https://doi.org/10.21432/T26W2K

Karnell, M. P., Schultz, K. et Canady, J. (2001). Investigations of a pressure-sensitive theory of marginal velopharyngeal inadequacy. *The Cleft Palate Craniofacial Journal, 38*(4), 346-357. https://doi.org/10.1597/1545-1569_2001_038_0346_ioapst_2.0.co_2

Klintö, K., Salameh, E.-K., Svensson, H. et Lohmander, A. (2011). The impact of speech material on speech judgement in children with and without cleft palate. *International Journal of Language & Communication Disorders, 46*(3), 348-360. https://doi.org/10.3109/13682822.2010.507615

Kuehn, D. P. et Moller, K. T. (2000). Speech and language issues in the cleft palate population: The state of the art. *The Cleft Palate Craniofacial Journal, 37*(4), 1-35. https://doi.org/10.1597/1545-1569_2000_037_0348_saliit_2.3.co_2

Kuehn, D. P. et Moon, J. B. (1998). Velopharyngeal closure force and levator veli palatini activation levels in varying phonetic contexts. *Journal of Speech, Language, and Hearing Research, 41*(1), 51-62. https://doi.org/10.1044/jslhr.4101.51

Kummer, A. W. (2011a). Perceptual assessment of resonance and velopharyngeal function. *Seminars in Speech and Language, 32*(2), 159-167. https://doi.org/10.1055/s-0031-1277718

Kummer, A. W. (2011b). Types and causes of velopharyngeal dysfunction. *Seminars in Speech and Language, 32*(2), 150-158. https://doi.org/10.1055/s-0031-1277717

Kummer, A. W. (dir.). (2020). Speech/resonance disorders and velopharyngeal dysfunction. Dans *Cleft palate and craniofacial conditions: A comprehensive guide to clinical management* (4e éd., p. 259-293). Jones & Bartlett Learning.

Lee, A., Potts, S. et Bressmann, T. (2020). Speech-language therapy students' auditory-perceptual judgements of simulated concurrent hypernasality and articulation disorders. *Clinical Linguistics & Phonetics, 34*(5), 479-492. https://doi.org/10.1080/02699206.2019.1655666

Lewis, K. E., Watterson, T. et Quint, T. (2000). The effect of vowels on nasalance scores. *The Cleft Palate Craniofacial Journal, 37*(6), 584-589. https://doi.org/10.1597/1545-1569_2000_037_0584_teovon_2.0.co_2

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research, 35*(6), 382-386. https://doi.org/10.1097/00006199-198611000-00017

Marino, V. C. C., Dutka, J. C. R., Manicardi, F. T., Gifalli, G., Silva, P. P. et Pegoraro-Krook, M. I. (2020). Influence of speech stimuli in the auditory perceptual identification of hypernasality in individuals with cleft lip and palate. *CoDAS, 32*(6), e20190269. https://doi.org/10.1590/2317-1782/20202019269

Martin, P., Beaudoin-Bégin, A.-M., Goulet, M.-J. et Roy, J.-P. (2001). Les voyelles nasales en français du Québec. *La linguistique, 37*, 49-70. https://doi.org/10.3917/ling.372.0049

Martin, P. (1996). *Éléments de phonétique avec application au français.* Les Presses de l'Université Laval.

Moon, J. B., Kuehn, D. P. et Huisman, J. J. (1994). Measurement of velopharyngeal closure force during vowel production. *The Cleft Palate Craniofacial Journal, 31*(5), 356-363. https://doi.org/10.1597/1545-1569_1994_031_0356_movcfd_2.3.co_2

Ostiguy, L. et Tousignant, C. (2008). *Les prononciations du français québécois : Normes et usages* (2e éd.). Guérin universitaire.

Peterson-Falzone, S. J., Trost-Cardamone, J., Karnell, M. P. et Hardin-Jones, M. A. (dir.). (2017). Effects of cleft and noncleft VPI on speech. Dans *The clinician's guide to treating cleft palate speech* (2e éd., p. 56-87). Elsevier Health Sciences.

Pluye, P. (2019). L'intégration en méthodes mixtes. Dans V. Ridde et C. Dagenais (dir.), *Évaluation des interventions de santé mondiale : Méthodes avancées* (p.187-212). IRD Éditions.

Polit, D. F., Beck, C. T. et Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459-467. https://doi.org/10.1002/nur.20199

Reinke, K. et Ostiguy, L. (2016). *Le français québécois d'aujourd'hui.* De Gruyter. https://doi.org/10.1515/9783110349306

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S. et Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*(2), 94-104. https://doi.org/10.1093/swr/27.2.94

Sell, D. (2005). Issues in perceptual speech analysis in cleft palate and related disorders: A review. *International Journal of Language & Communication Disorders, 40*(2), 103-121. https://doi.org/10.1080/13682820400016522

Todic, J., Sanguinetti, K. et Leuchter, I. (2022). Normative nasalance values in a population of French- speaking children. *Canadian Journal of Speech-Language Pathology and Audiology, 46*(3), 161-170. https://www.cjslpa.ca/files/2022_CJSLPA_Vol_46/No_3/CJSLPA_Vol_46_No_3_2022_1245.pdf

Tourmel, M. (2012). Établissement de normes de nasalance sur une population adulte de langue maternelle française ne présentant pas de rhinolalie [mémoire de maîtrise, Université de Genève]. Archive ouverte UNIGE. http://archive-ouverte.unige.ch/unige:23484

Trost-Cardamone, J. E. (2012, avril). *American English Sentence Sample: A controlled sample for assessing cleft palate speech outcome* [communication par affiche]. American Cleft Palate-Craniofacial Association Annual Meeting, San Jose, CA, USA.

Walter, H. (1983). La nasale vélaire/ŋ/ un phonème du français? *Langue française, 60*, 14-29. https://www.jstor.org/stable/41558161

Wouters, R., De Fraine, B. et Simons, M. (2019). What is at stake in deliberative inquiry? A review about a deliberative practice. *Systemic Practice and Action Research, 32*, 193-217. https://doi.org/10.1007/s11213-018-9457-8

Zajac, D. J. et Linville, R. N. (1989). Voice perturbations of children with perceived nasality and hoarseness. *The Cleft Palate Journal, 26*(3), 226-232. http://cleftpalatejournal.pitt.edu/ojs/cleftpalate/article/view/1277/1277

Note des auteures

Les demandes au sujet de cet article doivent être adressées à Marie-Ève Caty, Département d'orthophonie, Université du Québec à Trois-Rivières, 3351 boulevard des Forges, C.P. 500, Trois-Rivières, QC, Canada, G9A 5H7. Courriel : **marie-eve.caty@uqtr.ca**

La liste de phrases pour l'évaluation auditivo-perceptive de la résonance et de la parole en lien avec la fonction vélopharyngée en français québécois ainsi que les images et le manuel de passation sont disponibles gratuitement à l'adresse suivante : http://www.uqtr.ca/orthophonie/materiel.

Mme Eugénie Préfontaine est maintenant affiliée au département d'orthophonie, Centre hospitalier de l'Université de Montréal, QC, CANADA.

Remerciements

Ce projet a été rendu possible grâce au soutien des trois centres hospitaliers pédiatriques et à la communauté de pratique en dysfonction vélopharyngée et malformations orofaciales. Merci aux personnes qui ont contribué à la collecte de données : Julie Gauthier-Desgagné, Maryse Lavoie, Caroline Richer et Marie-Noël Malo.

Déclaration

Les auteures déclarent n'avoir aucun conflit d'intérêts, financiers ou autres.

The Impact of the Pandemic on Development: Parents' Perceptions on Language and Literacy

Les conséquences de la pandémie sur le développement des enfants : perceptions parentales sur le langage et la litéracie

Émilie Courteau
Guillaume Loignon
Mélanie Dutemple
S. Hélène Deacon

KEYWORDS
LANGUAGE
LITERACY
PANDEMIC
PARENTAL PERCEPTION
SOCIAL SKILLS

Émilie Courteau[1],
Guillaume Loignon[2],
Mélanie Dutemple[1,3],
S. Hélène Deacon[1]

[1]Department of Psychology & Neuroscience, Dalhousie University, Halifax, NS, CANADA

[2]Département d'éducation et pédagogie, Faculté des sciences de l'éducation, Université du Québec à Montréal, Montréal, QC, CANADA

[3]Département des sciences de l'éducation, Université du Québec en Outaouais, St-Jérôme, QC, CANADA

Editor:
Stefano Rezzonico

Abstract

This exploratory research examined how Canadian parents perceived the impact of the COVID-19 pandemic on several aspects of their children's development, and whether these perceptions differed between parents of children with and without developmental difficulties. Pandemic-related restrictions limited social interactions, likely leading parents to have concerns about the impact of the pandemic on children's development of social skills. However, it was not clear if parents were concerned about the potential impact of the pandemic on other areas of child development, such as language and early literacy. To examine parents' perceptions of the impact of the COVID-19 pandemic on child development, we conducted an online survey with 253 parents of preschool-aged children. Survey items covered two domains: possible impacts of pandemic-related measures and concerns about specific developmental skills. Parents reported that the most negative impact on their children's development during the pandemic was limitations on playing with other children. They were particularly concerned about the impact on social skills, more so than on language and early literacy skills. The results also showed that parents who suspected that their children had a developmental difficulty were more concerned about the pandemic's impact on their children's development than parents of children with a diagnosis or no difficulties. We discuss the implications of these findings considering current research on the effects of the COVID-19 pandemic on children's language, early literacy, and social skill development.

Abrégé

Cette recherche exploratoire porte sur la perception des parents quant aux conséquences de la pandémie de COVID-19 sur le développement de leurs enfants et vise à démontrer si ces perceptions diffèrent entre les parents d'enfants présentant ou non des difficultés de développement. Le manque d'interactions sociales découlant des restrictions liées à la pandémie pourrait être une préoccupation majeure puisqu'il pourrait être responsable de certaines répercussions sur le développement des compétences sociales. La préoccupation des parents concernant les conséquences potentielles de la pandémie sur d'autres domaines du développement, tels que le langage et la littéracie, n'est cependant pas établie. Nous avons mené une enquête en ligne à laquelle ont répondu 253 parents d'enfants d'âge préscolaire. Les questions portaient sur deux domaines : les effets de diverses mesures liées à la pandémie et les préoccupations concernant les compétences développementales. Les parents ont indiqué que l'effet le plus négatif sur le développement de leurs enfants durant la pandémie est la limitation liée aux jeux avec d'autres enfants. Les parents étaient particulièrement préoccupés par les conséquences sur les habiletés sociales. Les résultats montrent également que les parents soupçonnant une difficulté de développement chez leurs enfants étaient plus préoccupés par les effets de la pandémie que les parents d'enfants ayant déjà reçu un diagnostic ou n'ayant pas de difficultés développementales. Nous discutons des résultats obtenus à la lumière des recherches actuelles sur les effets de la pandémie de COVID-19 sur le développement du langage et de la littéracie et sur les compétences sociales des enfants.

During the COVID-19 outbreak, there were many community-based public health measures (hereafter referred to as "pandemic-related measures"), including mandated mask-wearing and social distancing. These measures have been important for disease prevention, but possible unintended consequences on child development might have caused concern among parents. Canadian mothers of children aged 0 to 5 years cited the lack of social interactions as one of their top concerns about pandemic-related measures (Sanders et al., 2022). This finding suggested that parents were preoccupied about the impact of the pandemic on the development of their children's social skills. However, it was not clear if parents were concerned about the pandemic's potential impacts on other developmental areas, such as language and early literacy. Parents' perceptions of these areas are important in part because parents' concerns about language development are essential to connecting children with the supports they need, such as diagnoses by professionals (Doove et al., 2021). This gap in understanding motivated us to examine parents' perceptions of the impact of the pandemic on the development of social, language, and early literacy skills in preschool-aged children. We did so for typically developing children and those with developmental delays. Emerging evidence suggests that children with developmental disabilities were particularly vulnerable to the pandemic's effects because they rely to a great extent on health care and community services (Aishworiya & Kang, 2021) that were likely to have been disrupted by the pandemic. This exploratory research surveyed Canadian parents' perceptions of the impact of pandemic-related measures on their children's language, early literacy, and social skills development.

Pandemic-Related Measures and Impacts on Child Development

Numerous pandemic-related measures were implemented globally to control the spread of COVID-19, including restrictions on gatherings (Kaplan et al., 2020), meaning that many children were unable to play with their friends. Many extracurricular activities were also cancelled (Ontario COVID-19 Science Advisory Table, 2021). Schools and daycares were closed for extended periods around the world, including here in Canada. Canadian speech-language pathologists (S-LPs) were advised to engage in virtual practice during the pandemic (see Wong, 2020, for a review). And even when in-person activities were permitted, face masks were mandated, including in daycares and schools (Jackson, 2021). It would be reasonable to expect that such measures may have an impact on the development of social, language, and literacy skills of young children.

Pandemic-related measures have been demonstrated to have negative effects on children's social interactions. A meta-analysis by Viner and colleagues (2022) found that school closures and social restrictions were associated with an increased prevalence of mental health and emotional problems in children from preschool to adolescence, including increased anxiety and depressive symptoms. López-Bueno and colleagues (2021) suggested that social isolation during lockdowns may have led to increases in negative behaviours in preschool and school-aged children due to insufficient physical activity and excessive screen exposure.

There is less evidence of pandemic impacts on language development, and the direction of these effects is unclear. On one hand, there are several ways in which pandemic-related measures may have had negative impacts on language development. Extensive mask-wearing in a child's environment, such as daycare, may have made it more difficult to use lipreading for language learning, considering that infants as young as 8 months of age use information from lips as a part of their learning of native speech forms (Lewkowicz & Hansen-Tift, 2012). Indeed, mask-wearing in response to the pandemic has been associated with poor speech recognition for 4- and 5-year-old children, although not for older children (Kwon & Yang, 2023). Language development may have been negatively affected by disruptions in health services, such as the suspension of universal newborn hearing screening (e.g., July 2020 in Ontario). Early detection and treatment of congenital hearing loss are crucial for the development of speech and language (Runnion & Gray, 2019). On the other hand, it is also possible that pandemic-related measures might have had positive effects on language development. Based on normative data, Kartushina et al. (2022) found that children aged 8 to 36 months made larger vocabulary gains than expected during the daycare closures; this growth was observed in children who had less exposure to screens (e.g., watching television) and who were read to regularly and frequently. Therefore, pandemic-related measures may have had positive effects on language development in that caregivers had more time to talk and read with their children, in some cases offsetting negative effects.

Similarly, negative and positive effects of pandemic-related measures might emerge for early literacy skills. A longitudinal study by Roy (2022) found a decline in word reading skills between Grade 1 and 2 for children tested during the first 2 years of the pandemic. These findings were consistent with Bao et al.'s (2020) prediction that kindergarteners' reading ability would decline by 66% due to school closures. However, parents with children aged 2

to 4 years reported that they read more to children during COVID-19 compared to before, in part because they were home more often with their young children during lockdowns (Wheeler & Hill, 2021). Thus, like what has been reported for language skill development, some negative effects of pandemic-related measures may have been offset due to increased early literacy stimulation by more available parents.

Emerging evidence suggests that the pandemic had a more negative impact on children with developmental disabilities, who comprise 6.5% to 8.3% of all children in Canada (Berrigan et al., 2023), than on children with normal development (Aishworiya & Kang, 2021). This is in part due to interruptions or changes in health and support services (Aishworiya & Kang, 2021). Tohidast et al. (2020) explained that S-LP services were interrupted or delayed, possibly at critical stages of speech and language development; these disruptions in treatment have caused multiple problems for children and their families during COVID-19 lockdowns. For example, Tohidast et al. (2020) suggested that the pandemic may have adversely affected the quality of parental care and the way parents conducted speech-language exercises due to pandemic-related stress.

Parents of children with autism spectrum disorder (ASD) reported concerns about functional, social, and behavioural changes in their children, including a lack of communication opportunities due to service closures and social restrictions (Tokatly Latzer et al., 2021). Other effects have emerged for children with behavioural difficulties, such as inattention and hyperactivity. Wendel et al. (2020) found an increase in parent reports of attention-deficit/hyperactivity disorder (ADHD) symptoms in 4- and 5-year-old Canadian children before and after the onset of the COVID-19 pandemic, possibly linked to kindergarten virtual learning.

Overall, the pandemic disrupted the lives of Canadian families. Sanders et al. (2022) interviewed 10 Canadian mothers with preschool-aged children about how the pandemic affected their daily lives. The families cited the disruption of services as a major impact of the pandemic, leaving them exhausted and worried about their children's development.

The Present Study

The effects of pandemic-related measures reported in the literature were mostly negative for children's development. However, some positive effects were reported, related to parents having more time to stimulate their children by being at home with them more. Parents of children with developmental difficulties may have also experienced greater negative effects. The purpose of this exploratory study was to learn more about parental perceptions and the pandemic by addressing three research questions.

1. How do Canadian parents of preschool children (0 to 5 years) perceive the impact of the pandemic on their children's development, considering five pandemic-related measures that likely affected their lives?
2. What are parents' perceptions regarding the pandemic's effects on specific aspects of their children's development, such as social skills, language, and early literacy skills?
3. How do parents' perceptions differ between those who have children with developmental difficulties, those who have children without, and those who express concerns about their children's development without a diagnosis?

Both clinicians and public health measures benefit from knowing more about parents' perceptions, as this can help improve support systems and communication strategies. In addition, parents' concerns are also related to seeking early intervention for their child if needed (Doove et al., 2021). We conducted an online survey with questions assessing the parents' perceptions in each of two domains: pandemic-related measures and their children's skill development. As the first step of a validation process, we applied a clustering method to determine if the questions we designed aligned with our two domains of interest.

Method

Participants

A total of 253 participants were recruited for the study using Qualtrics (https://www.qualtrics.com), a commercial survey platform. Qualtrics contacted potential respondents who met our eligibility criteria from a pool of people who had previously expressed interest in participating in surveys. To be eligible for the study, participants had to be 18 years or older, reside in Canada, and have a child aged 5 years or younger. The participants and survey items analyzed in this study are part of a larger research project called the *Language and Literacy Environment Questionnaire*, which aimed to investigate how parents support their children's language and literacy skills at home, both in reading on paper and on screen. The protocol was approved by the Research Ethics Board of Dalhousie University (REB #2021-5570). Data were collected in June and July of 2022. Sixteen respondents were removed from the dataset as, after verification, they did not meet the eligibility criteria. Data from the remaining 237 respondents were analyzed. Their demographic characteristics are summarized in **Table 1**.

Table 1

Distribution of Respondents by Demographic Characteristics

	Number of respondents (N = 237)	%
Child's age (months)		
0–11	35	14.8
12–23	52	21.9
24–35	37	15.6
36–47	30	12.7
48–59	40	16.9
60–71	43	18.1
Child's first language		
English	200	84.4
French	20	8.4
Cantonese	5	2.1
German	1	0.4
Other	11	4.6
Province or territory		
Atlantic provinces	16	6.8
Québec	21	8.9
Ontario	109	46.0
Manitoba	14	5.9
Saskatchewan	8	3.4
Alberta	37	15.6
British Columbia	32	13.5
Highest level of education		
Some high school (Grade 9 or higher)	4	1.7
High school graduate	38	16.0
Some college/university	29	12.2
College/university graduate	121	51.1
Some postgraduate	10	4.2
Postgraduate degree	35	14.8

Note. Percentage calculated on the total number of respondents, N = 237.

We divided the respondents in three groups based on the following survey question: "Does your child have any difficulties in the following areas?" The areas were speech or language, hearing, ASD, reading, learning, behaviour, ADHD, and others. The response options were: "NO, "MAYBE but not diagnosed," and "YES - diagnosed by a professional." Respondents who answered "no" to all the development areas listed formed the NO group (n = 168). Respondents who answered "yes" to at least one developmental area formed the YES group (n = 23). Respondents who answered "maybe" at least once and did not answer "yes" to any areas listed formed the MAYBE group (n = 46). No difference was found between groups for the child's gender ($X^2(2) = 0.117$, $p = .943$) and parental education ($X^2(8) = 7.43$, $p = .47$). Differences between groups were found in age, with the NO group having a significantly lower mean age (in months; $M = 32.7$, $SD = 20.4$) compared to the "MAYBE" group ($M = 42.7$, $SD = 21.4$, $p = .012$) and the "YES" group ($M = 43.5$, $SD = 21.1$, $p = .029$).

Procedure

The entire questionnaire, along with the consent form, was written in English, and completed online using the Qualtrics survey software. Respondents were informed that completing the questionnaire would take approximately 20 to 30 min and that their participation was completely voluntary and anonymous. Upon completion of the questionnaire, respondents were compensated via Qualtrics recruitment services.

The questionnaire underwent a clear language and design revision by a hired professional. This allowed us to ensure that the questionnaire was written in clear or simple language so that respondents could easily understand the message and respond accordingly, making it easier for people with lower literacy skills to access information. The sections of the questionnaire relevant to the present study are described below.

Demographic Information

Parents answered demographic questions about themselves (e.g., marital status, education), their family (e.g., household income), and their youngest child of 5 years old or younger (e.g., age, gender).

Parent's Perception of the Pandemic

Given the lack of questionnaires on parents' perception of the pandemic and its effect on their children's development, we developed two question sets: one pertaining to pandemic-related measures and the other on the development of children's skills. Each set was comprised of a prompt, followed by a list of items (pandemic-related measures or skills) that respondents rated on a 5-point Likert scale in line with Dillman et al.'s (2014) recommendations: *Very good; Somewhat good; Neither good nor bad; Somewhat bad; Very bad*.

The first survey questions were related to pandemic-related measures and used the following prompt: "Do you think the following activities were good or bad for your child's development during the pandemic, and how much?" It contained five items displayed at once, one on each of the following pandemic-related measures: mask-wearing (by your child or others), daycare or school closures, virtual learning, cancellation of in-person extracurricular activities, and not being able to play with other kids in person. The second set of questions was related to skill development and used the following prompt: "Do you think the COVID-19 pandemic has been good or bad for your child's skills in the following areas and how much?" It contained three items: language skills, reading skills, and social skills.

Analyses

Because the survey questions were developed for this study, we used a partitioning method to verify if the items aligned in a meaningful and coherent way with the other items in their question set (Sireci & Geisinger, 1992), that is, pandemic-related measures or skill development. This verification was performed through Ward's method of hierarchical clustering (Murtagh & Legendre, 2014) on Spearman rank correlations.[1] To assess the parents' perception of pandemic-related measures and their impacts on their children's skill development, we performed separate analyses of variance (type II analysis of variance [ANOVA], using a significance level of .05) on the two question sets. Parents' perceptions were evaluated based on their responses to 5-point Likert scale items. The first ANOVA included the factor Group (YES, MAYBE, and NO) and Measure (mask-wearing by children of others, daycare or school closures, virtual learning, cancellation of in-person activities and not being able to play with friends in person). The second ANOVA included the factor Group (YES, MAYBE, and NO) and Skill (language, reading, and social).[2] In both ANOVAs, children's age and region[3] were also included to account for their potential influence on parents' perceptions of the pandemic. Post hoc analyses were conducted using the marginal effects method (Lenth, 2023), enabling comparisons between levels of a variable while accounting for the other factors in the model; results are reported with 95% confidence intervals and Cohen's *d* effect sizes obtained through the marginal effects method. Post hoc *p* values were adjusted using the Benjamini-Hochberg procedure.

Results

Survey Item Clustering

The hierarchical survey item clustering procedure resulted in two main clusters aligned with the theorized question sets, that is, pandemic-related measures and

[1] Ward's method implemented in the hclust function of the R language (The R Foundation, 2022).
[2] Plots of the residuals revealed no severe departure from normality.
[3] We included age to account for differences between groups and the variable region because pandemic-related measures may have differed by region in Canada. We refer to province or territory as region; see **Table 1**. To ensure that each level of the region variable had sufficient observations to run an ANOVA analysis, we combined two of the prairie provinces (Saskatchewan and Manitoba).

developmental skills, as shown in **Figure 1**. Meaningful subclusters also appeared. Within the pandemic-related measures, there was a subcluster composed of the two items about prohibiting personal interactions (i.e., "Not being able to play with other kids in-person" and "Cancellation of in-person extracurricular activities"). Within the developmental skill question set, a second subcluster emerged, capturing the reading skill and language skill items.

Pandemic-Related Measures

Parents' perceptions varied across the five pandemic-related measures items. **Figure 2** suggests that for most measures, the MAYBE group had the most negative view of the impact of the pandemic-related measures on the children's development compared to the YES and NO groups.

Significant main effects of the two-way ANOVA indicated that the reported perception differed by Group ($F(2) = 5.519$, $p < .001$) and Measure ($F(4) = 15.507$, $p < .001$). The interaction between Group and Measure was not statistically significant, $F(8) = 0.547$, $p = .822$. Age ($F(1) = 1.886$, $p = 0.170$) and Region ($F(5) = 1.3998$, $p = 0.222$) were also included in the model but did not show statistically significant main effects.

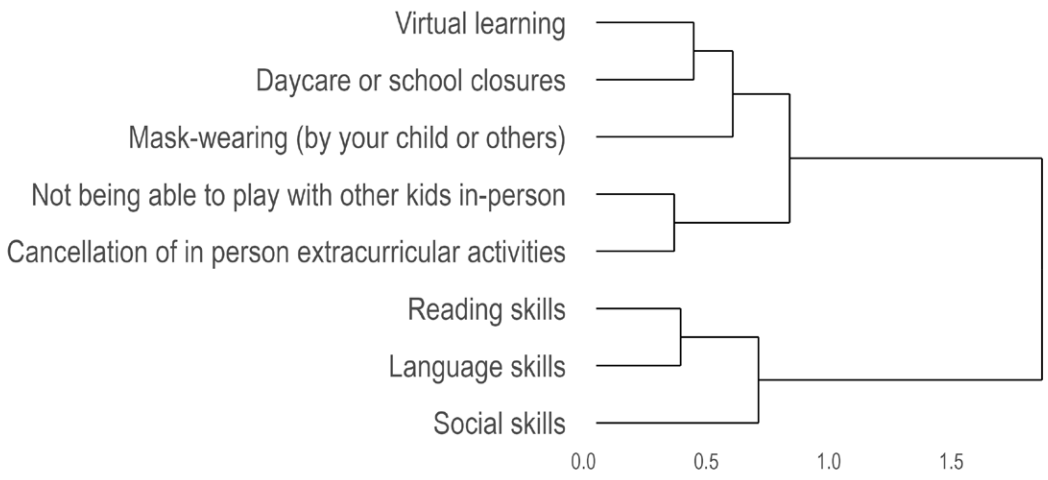
Post hoc pairwise comparisons of the Group factor revealed that when examining the impact of pandemic-related measures on their children's development, the MAYBE group ($M = 3.61$, 95% CI [3.45, 3.78]) reported a more negative view compared to the YES group ($M = 3.29$, 95% CI [3.06, 3.52]), $p = .028$, $d = 0.28$. The MAYBE group also had a more negative view than the NO group ($M = 3.33$, 95% CI [3.24, 3.42]), $p = .004$, $d = 0.24$. No significant difference was found between the YES group and the NO group, $p = .740$.

Post hoc analyses were conducted on the Measure factor, comparing each measure with the remaining measures combined. Parents reported a significantly more negative impact for "Not being able to play with friends in person" ($M = 3.81$, 95% CI [3.60, 4.02]) compared to the other pandemic-related measures, $p < .001$, $d = 0.42$. In contrast, parents perceived "Virtual learning" ($M = 3.06$, 95% CI [2.85, 3.27]) as having a less negative effect than the other pandemic-related measures, $p < .001$, $d = -0.38$. The perception of "mask-wearing" ($M = 3.22$, 95% CI [3.01, 3.42]) showed a notable trend of being less negative compared to the other measures, $p = .055$, $d = -0.21$. Other measures did not show significant differences.

Developmental Skills

Visual inspection of **Figure 3** suggests that, as with pandemic-related measures, the MAYBE group generally perceived the impact of the pandemic on skill development as more negative than the YES or NO groups. **Figure 3** also suggests that respondents believed that COVID-19 had a more negative impact on the development of their children's social skills than on language and literacy skills. This was supported by the statistical analysis.

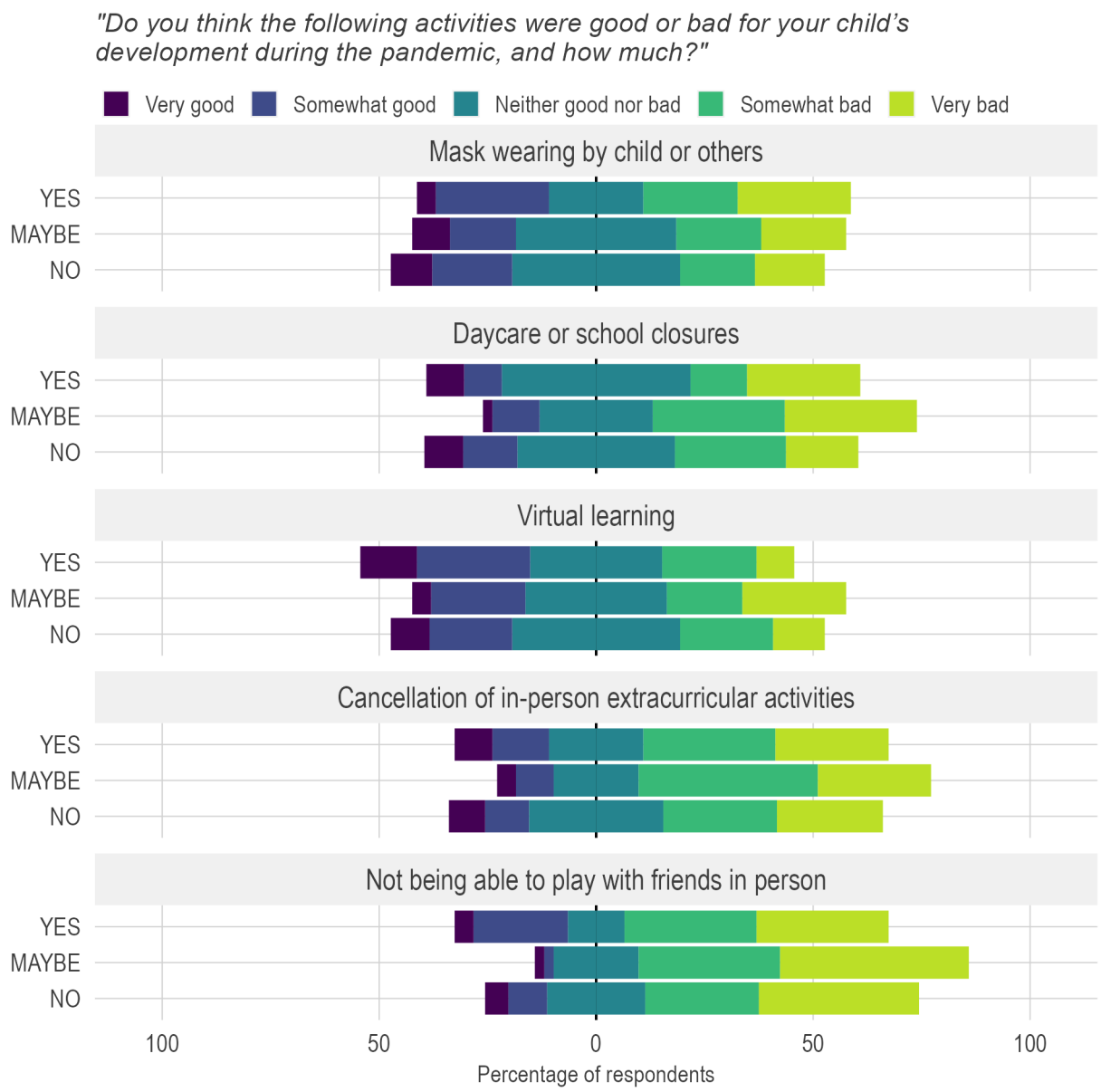
Figure 1



Hierarchical clustering of the questions included in the survey

Note. Ward hierarchical clustering using Spearman rank correlations. The x axis is the value of the Ward minimum variance criterion, used as a distance measurement.

Figure 2



Parents' perceptions of pandemic-related measures

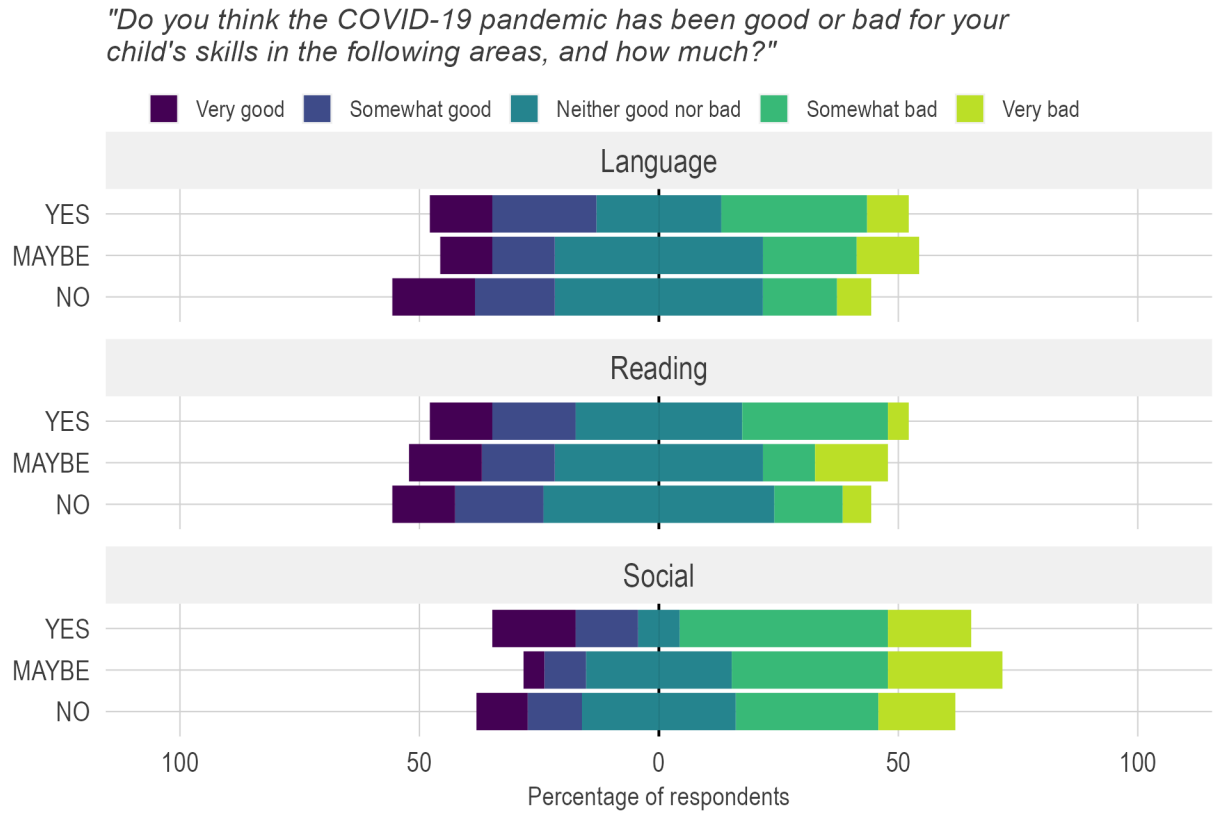
Note. Distribution of Likert-scale responses across 5 levels (*Very good* to *Very bad*), by pandemic-related measure and parent-reported presence of child developmental difficulty (YES, MAYBE, NO). $N = 237$.

ANOVA analysis revealed significant main effects for Group, $F(2) = 3.025$, $p = .049$, and Skill, $F(3) = 15.165$, $p < .001$. The interactions between Group and Skill, $F(4) = 0.268$, $p = .899$; Age, $F(1) = 0.395$, $p = .530$; and Region, $F(5) = 1.792$, $p = .110$, were not statistically significant..

Post hoc pairwise comparisons of groups showed that the MAYBE group had a significantly more negative perception ($M = 3.21$, 95% CI [3.01, 3.42]) compared to the NO group ($M = 2.94$, 95% CI [2.82, 3.05]), $p = .042$. There was no significant difference in perception between the YES group ($M = 3.01$, 95% CI [2.72, 3.29]) and the NO group, $p = .662$, or between the YES group and the MAYBE group, $p = .346$.

Post hoc comparisons were conducted to examine the perceived impact of the pandemic on each skill, compared

Figure 3



Parents' perceptions of the impact of COVID-19 on children's developmental skills

Note. Distribution of Likert-scale responses across 5 levels (*Very good* to *Very bad*), by developmental skill and parent-reported presence of child developmental difficulty (YES, MAYBE, NO). $N = 237$.

to other two skills combined. Results showed that parents reported a more negative impact for social skills ($M = 3.37$, 95% CI [3.16, 3.58]) compared to other skills combined, $p < .001$, $d = 0.42$. In contrast, parents reported a significantly less negative impact of the pandemic on reading skills ($M = 2.87$, 95% CI [2.66, 3.08]), $p = .037$, $d = -0.25$. The perceived impact on language skills ($M = 2.92$, 95% CI [2.71, 3.13]) did not significantly differ from the other skills, $p = .115$.

Discussion

The purpose of this exploratory study was to understand parents' perceptions of the impact of the pandemic on their preschool-aged children's development by examining their views on five specific pandemic-related measures and how the pandemic affected their children's language, early literacy, and social skills. Additionally, the study examined differences in parents' perceptions between those who reported having a child with developmental difficulties and those who did not, as well as those who reported concerns about their children's development without a formal diagnosis.

Validity evidence for our survey came from clustering analyses. These analyses showed that the survey items were organized into the hypothesized distinct domains (i.e., pandemic-related measures and developmental skills), with some subclusters in which parents' perceptions of language and reading skills were correlated in a subcluster separate from social skills.

Our first research questions addressed how parents reported the impact of five pandemic-related measures on their children's development. We present the results from the least perceived negative impact to the most impactful in the parents' view. We found that parents perceived "virtual learning" as having the least negative impact on their children's development, with many parents rating it as very good or somewhat good (see **Figure 2**). It is possible that parents perceived the virtual sessions as more positive than the complete absence of school or daycare. Another interpretation could be that parents were less likely to be concerned about virtual learning because this measure was less likely to target preschool children. Nevertheless, this

finding is of interest to clinicians, including S-LPs, because it suggests that many parents were receptive to virtual learning and may also be open to virtual therapy.

The finding that "mask-wearing" tended to be a measure that parents were less concerned about is noteworthy, particularly considering the prominent and heated public debates on the issue. One interpretation of this finding is that parents were aware of the essential role of wearing a mask in preventing the spread of COVID-19, thus seeing it as a positive measure for their children's health and development. During the pandemic, it is likely that parents had to negotiate between wearing masks to promote their family's health and the potential negative effects of masks on language processing in children (e.g., Kwon & Yang, 2023). It would be useful for clinicians to hold discussions with parents and educators and provide them with strategies for enhancing communication while wearing masks, like ensuring that the child's attention is focused on you before speaking or using gestures to support communication (see Baltimore & Atcherson, 2020, for more recommendations).

When compared to other measures, parents reported that "Not being able to play with other kids in-person" was the most negative for their children's development, which limited their children's social interactions. This finding is consistent with previous research indicating that parents frequently cited the impact of the pandemic on their children's social skills as a primary concern (e.g., Tokatly Latzer et al., 2021; Sanders et al., 2022). In addition, this result is in line with the parents' perception of the impact of the pandemic on developmental skills. In relation with our second research question, we found that parents perceived greater negative impacts of the pandemic on their children's social skills than on their language and early literacy skills. These results are not surprising, as the impact of the pandemic on children's social skills was probably the most discussed issue during the pandemic. One question that arises from our exploratory study's results is whether parents are informed about the role of language acquisition in the development of social skills. Doove et al. (2021) found a direct association between language development and social competence in preschoolers. A potential avenue for further research is to assess whether parents and educators are knowledgeable that children's social competence is related to language development. This would be valuable information because parents' and educators' concerns are related to early detection of language deficits (Doove et al., 2021).

The third research question aimed to determine whether parents' perceptions differed according to whether they reported having a child with or without developmental difficulties. The results indicated that parents who suspected that their child had at least one developmental difficulty (the MAYBE group) reported higher levels of negative impacts of the pandemic than parents reporting having a child with or without difficulties (YES and NO groups), who reported similar perceptions. This pattern was somewhat surprising to us, and yet, we think that it suggests an association between uncertainty and perception of the pandemic's impact. Sanders et al. (2022) listed gaps in health care as one of parents' primary concerns, with very clear evidence of shifts in health services across the country. Parents suspecting that their children had difficulties were likely to have felt frustrated in their attempts to get services, likely waiting even longer than normal for diagnostic and treatment services. These concerns would be justified, as parents' concerns are often followed by an actual diagnosis (Doove et al., 2021). Delays might have been even longer for those without a diagnosis who are not yet connected in the system. The absence of a diagnosis could also mean that parents do not have social supports such as through a support group, which may increase negative perceptions of the impact of the pandemic, with its many other restrictions on social connections. As a result, parents had to take responsibility for finding, filtering, understanding, and then using information to make health-related decisions about their children (Doove et al., 2021). Although these findings are suggestive, we cannot rule out the possibility that parents reporting that their children might have a developmental disability not yet diagnosed may include respondents who are generally more anxious about their children's development. They may also have had a more negative view of general pandemic prevention measures. This is a limitation of our study as these dimensions were not included in our survey.

The equivalent level of concern reported by the parents with and without developmental disabilities is surprising, as parents of children with diagnosed difficulties were expected to be more affected by the pandemic than parents with typically developing children (Aishworiya & Kang, 2021). A first interpretation is that for children with diagnosed difficulties many health care services, such as S-LP therapy, were available virtually during the pandemic (Wong, 2020). One limitation of our study is that we did not ask specifically about the cancellation of health services as a pandemic-related measure. A second explanation may arise from the positive effects of the pandemic on language and early literacy skills (Kartushina et al., 2022). Being more often at home with their children, parents may have had more time for developmentally beneficial activities, such as reading with their children (Wheeler & Hill, 2021).

In conclusion, our exploratory research provided insight into how parents perceived the pandemic's impact on their children's development. The results suggest that social skills development is a main preoccupation for Canadian parents. Awareness should be raised that, although parental concern about social skills is legitimate, these skills are closely linked to preschool children's language development.

References

Aishworiya, R., & Kang, Y. Q. (2021). Including children with developmental disabilities in the equation during this COVID-19 pandemic. *Journal of Autism and Developmental Disorders, 51*(6), 2155–2158. https://doi.org/10.1007/s10803-020-04670-6

Baltimore, W. J., & Atcherson, S. R. (2020, June 5). *Helping our clients parse speech through masks during COVID-19*. The ASHA Leader. https://leader.pubs.asha.org/do/10.1044/leader.MIW.25062020.34/full/

Bao, X., Qu, H., Zhang, R., & Hogan, T. P. (2020). Modeling reading ability gain in kindergarten children during COVID-19 school closures. *International Journal of Environmental Research and Public Health, 17*(17), Article 6371. https://doi.org/10.3390/ijerph17176371

Berrigan, P., Scott, C. W. M., & Zwicker, J. D. (2023). Employment, education and income for Canadians with developmental disability: Analysis from the 2017 Canadian survey on disability. *Journal of Autism and Developmental Disorders, 53*, 580–592. https://doi.org/10.1007/s10803-020-04603-3

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail and mixed-mode surveys: The tailored design method* (4th ed.). John Wiley & Sons.

Doove, B. M., Feron, F. J. M., van Os, J., & Drukker, M. (2021). Preschool communication: Early identification of concerns about preschool language development and social participation. *Frontiers in Public Health, 8,* Article 546536. https://www.frontiersin.org/articles/10.3389/fpubh.2020.546536

Jackson, H. (2021, December 8). *Mask mandates in Canada: A province-by-province look at the rules and exemptions.* CTVNews. https://www.ctvnews.ca/health/coronavirus/mask-mandates-in-canada-a-province-by-province-look-at-the-rules-and-exemptions-1.5698936

Kaplan, J., Frias, L., & McFall-Johnsen, M. (2020, September 23). *Our ongoing list of how countries are reopening, and which ones remain under lockdown*. Business Insider. https://www.businessinsider.com/countries-on-lockdown-coronavirus-italy-2020-3

Kartushina, N., Mani, N., Aktan-Erciyes, A., Alaslani, K., Aldrich, N. J., Almohammadi, A., Alroqi, H., Anderson, L. M., Andonova, E., Aussems, S., Babineau, M., Barokova, M., Bergmann, C., Cashon, C., Custode, S., De Carvalho, A., Dimitrova, N., Dynak, A., Farah, R., … Mayor, J. (2022). COVID-19 first lockdown as a window into language acquisition: Associations between caregiver-child activities and vocabulary gains. *Language Development Research, 2*, 1–36. https://doi.org/10.34842/abym-xv34

Kwon, M., & Yang, W. (2023). Effects of face masks and acoustical environments on speech recognition by preschool children in an auralised classroom. *Applied Acoustics, 202*, Article 109149. https://doi.org/10.1016/j.apacoust.2022.109149

Lenth, R. (2023). *emmeans: Estimated marginal means, AKA least-squares means*. R package version 1.8.5. https://CRAN.R-project.org/package=emmeans

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences, 109*(5), 1431–1436. https://doi.org/10.1073/pnas.1114783109

López-Bueno, R., López-Sánchez, G. F., Casajús, J. A., Calatayud, J., Tully, M. A., & Smith, L. (2021). Potential health-related behaviors for pre-school and school-aged children during COVID-19 lockdown: A narrative review. *Preventive Medicine, 143*, Article 106349. https://doi.org/10.1016/j.ypmed.2020.106349

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification, 31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

Ontario COVID-19 Science Advisory Table. (2021, June 16). *COVID-19 and education disruption in Ontario: Emerging evidence on impacts.* https://doi.org/10.47326/ocsat.2021.02.34.1.0

The R Foundation. (2021). *The R project for statistical computing*. https://www.R-project.org/

Roy, K. (2022). *English speaking and English learning students' early literacy performance before and during the COVID-19 pandemic* [Doctoral dissertation, University of Oregon]. Scholars' Bank. https://scholarsbank.uoregon.edu/xmlui/handle/1794/27736

Runnion, E., & Gray, S. (2019). What clinicians need to know about early literacy development in children with hearing loss. *Language, Speech, and Hearing Services in Schools, 50*(1), 16–33. https://doi.org/10.1044/2018_LSHSS-18-0015

Sanders, C., Frank, T. J., Amyot, T., Cornish, K., Koopmans, E., Usipuik, M., Irving, L., & Pelletier, C. A. (2022). Day-to-day life during the COVID-19 pandemic: A longitudinal qualitative study with Canadian parents of young children. *Contemporary Issues in Early Childhood,* Advance online publication. https://doi.org/10.1177/14639491221115475

Sireci, S. G., & Geisinger, K. F. (1992). Analyzing test content using cluster analysis and multidimensional scaling. *Applied Psychological Measurement, 16*(1), 17–31. https://doi.org/10.1177/014662169201600102

Tohidast, S. A., Mansuri, B., Bagheri, R., & Azimi, H. (2020). Provision of speech-language pathology services for the treatment of speech and language disorders in children during the COVID-19 pandemic: Problems, concerns, and solutions. *International Journal of Pediatric Otorhinolaryngology, 138*, Article 110262. https://doi.org/10.1016/j.ijporl.2020.110262

Tokatly Latzer, I., Leitner, Y., & Karnieli-Miller, O. (2021). Core experiences of parents of children with autism during the COVID-19 pandemic lockdown. *Autism, 25*(4), 1047–1059. https://doi.org/10.1177/1362361320984317

Viner, R., Russell, S., Saulle, R., Croker, H., Stansfield, C., Packer, J., Nicholls, D., Goddings, A.-L., Bonell, C., Hudson, L., Hope, S., Ward, J., Schwalbe, N., Morgan, A., & Minozzi, S. (2022). School closures during social lockdown and mental health, health behaviors and well-being among children and adolescents during the first COVID-19 wave: A systematic review. *JAMA Pediatrics, 176*(4), 400–409. https://doi.org/10.1001/jamapediatrics.2021.5840

Wendel, M., Ritchie, T., Rogers, M. A., Ogg, J. A., Santuzzi, A. M., Shelleby, E. C., & Menter, K. (2020). The association between child ADHD symptoms and changes in parental involvement in kindergarten children's learning during COVID-19. *School Psychology Review, 49*(4), 466–479. https://doi.org/10.1080/2372966X.2020.1838233

Wheeler, D. L., & Hill, J. C. (2021). The impact of COVID-19 on early childhood reading practices. *Journal of Early Childhood Literacy*, Advance online publication. https://doi.org/10.1177/14687984211044187

Wong, J. C. (2020). Pandemic planning for hospital-based speech-language pathologists: Emerging lessons from coronavirus disease. *Canadian Journal of Speech-Language Pathology and Audiology, 44*(3), 151–165. https://cjslpa.ca/files/2020_CJSLPA_Vol_44/No_3/CJSLPA_Vol_44_No_3_2020_1236.pdf

Authors' Note

Correspondence concerning the article should be addressed to Émilie Courteau, Department of Psychology & Neuroscience, Dalhousie University, 1355 Oxford St., Halifax, NS, CANADA, B3H 4J1. **Email: emilie.courteau@dal.ca**

Acknowledgments

This research was funded by the SSHRC grant *Partnering to Empower Parents to Support Full Literacy for Children* to Deacon (892-2020-2027) and a Mitacs' Accelerate grant to Deacon and the Canadian Children's Literacy Foundation. We gratefully thank Nina Shukla for her help with this project.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Development, Validity, and Reliability of the Auditory and Speech Performance Test for Children

Développement, validité et fiabilité du *Auditory and Speech Performance Test for Children* (Test des habiletés auditives et langagières chez l'enfant)

Hale Hancer
Erhan Kiziltan
Pinar Civak Tan
Derya Gokmen
Serhat Hayme
Suna Tokgoz Yilmaz

KEYWORDS

AUDITORY PROCESSING
SPEECH PROCESSING
AUDITORY DISCRIMINATION
SPEECH DISCRIMINATION
REACTION TIME

Hale Hancer[1], Erhan Kiziltan[2], Pinar Civak Tan[3], Derya Gokmen[4], Serhat Hayme[5], Suna Tokgoz Yilmaz[6]

[1]Department of Audiology, Baskent University, Ankara, TÜRKIYE

[2]Department of Physiology, Baskent University, Ankara, TÜRKIYE

[3]Department of Audiology and Speech Disorders, Ankara University, Ankara, TÜRKIYE

[4]Department of Biostatistics, Ankara University, Ankara, TÜRKIYE

[5]Department of Biostatistics and Medical Informatics, Erzincan Binali Yildirim University, Erzincan, TÜRKIYE

[6]Department of Audiology, Ankara University, Ankara, TÜRKIYE

Editor:
Victoria Duda

Abstract

Auditory processing and speech processing disorders negatively affect school-aged children. To minimize these negative effects, individuals in the risk group should benefit from the positive contribution of early intervention with a comprehensive evaluation. The aim of this study was to develop the Auditory and Speech Performance Test for Children and analyze its validity and reliability. In the development of the Auditory and Speech Performance Test for Children, discrimination and recognition subtasks were built for both auditory and speech performance. Meaningful and meaningless minimal pairs were used in the subtasks. A silent background was used for auditory performance, and noise stimuli were combined into minimal pairs for speech performance. Audiovisual materials were integrated into the finger-tapping test. The Auditory and Speech Performance Test for Children was administered to 307 children with typical development and to 80 children with specific learning disabilities. The Auditory and Speech Performance Test for Children calculated children's reaction times for pressing speed and accuracy of pressing the correct key. The data were analyzed for content, construct validity, internal consistency, and test-retest reliability. The content validity index value was found to be high (.89–1.0). The Auditory and Speech Performance Test for Children was explained as a two-factor model using exploratory factor analysis (eigenvalue = 1.92, total variance = 66.65%). It was found to be discriminative according to age, groups, subtests, and 27% bottom and top scores (all were significant at $p < .001$). Internal consistency (.77–.90) and test-retest values (.89–.93) of the Auditory and Speech Performance Test for Children in the total test scores were calculated within reliable values. In conclusion, we developed a valid and reliable screening tool for auditory and speech performance in children.

Abrégé

Les troubles de traitement auditif et de traitement des informations langagières ont des effets négatifs sur les enfants d'âge scolaire. Afin d'atténuer ces effets négatifs, il serait préférable que les enfants à risque bénéficient d'une évaluation complète des habiletés de traitement auditif et de traitement des informations langagières, ainsi que des avantages que procure l'intervention précoce. L'objectif de cette étude était de développer le *Auditory and Speech Performance Test for Children (*Test des habiletés auditives et langagières chez l'enfant) et d'en mesurer la validité et la fiabilité. Pendant la phase de développement de ce test, des sous-tâches de discrimination et de reconnaissance ont été conçues pour évaluer les habiletés auditives et langagières. Des paires minimales composées de mots et de non-mots étaient utilisées dans ces sous-tâches. L'évaluation des habiletés auditives a été effectuée à l'aide de paires minimales sans la présence de bruit de fond, tandis que les habiletés langagières ont été évaluées à l'aide de paires minimales présentées avec un bruit de fond. Du matériel audiovisuel a été intégré à un test de tapotement du doigt. Le *Auditory and Speech Performance Test for Children* a été utilisé auprès de 307 enfants au développement typique et 80 enfants ayant un trouble spécifique des apprentissages. Il a mesuré le temps de réaction avec lequel les enfants appuyaient sur une touche et la précision des touches enfoncées. Les données ont été analysées de façon à déterminer les validités de contenu et de construit, la cohérence interne et la fiabilité test-retest. Les résultats ont montré que la valeur de l'indice de validité du contenu était élevée (0,89–1,0). Les résultats de l'analyse factorielle exploratoire ont montré que le *Auditory and Speech Performance Test for Children* était expliqué par un modèle à deux facteurs (valeur propre = 1,92, variance totale = 66,65 %). Les résultats ont aussi montré que le *Auditory and Speech Performance Test for Children* permettait de discriminer selon l'âge, le groupe, la sous-tâche et 27 % des scores supérieurs et inférieurs (tous étaient significatifs à $p < 0{,}001$). Les valeurs calculées pour la cohérence interne (0,77–0,90) et la fidélité test-retest (0,89–0,93) se retrouvaient à l'intérieur de l'intervalle de fiabilité. En conclusion, nous avons développé un outil de dépistage valide et fiable pour évaluer les habiletés auditives et langagières des enfants

Auditory processing disorder (APD) refers to impairment in the perceptual processing of auditory information in the central auditory nervous system (American Speech-Language- Hearing Association [ASHA], 2005a), whereas *speech processing disorder* refers to difficulty interpreting or comprehending auditory information (Richard, 2017). Because these disorders are not well known, they are often confused with other neurodevelopmental disorders such as specific learning difficulties (SLD), attention deficits, and hyperactivity disorders (ASHA, 2005a, 2005b; Bellis, 2011; Chermak et al., 1997).

A comprehensive assessment of auditory and speech processing skills is important for intervention and follow-up (ASHA, 2005a; Bellis, 2011). APD test batteries are ignored because they are not functional owing to the need for professional experts and ineffective use of time (ASHA, 2005b). There is currently no gold standard for evaluating speech or auditory processing skills (Palana et al., 2022; Richard, 2017). Consequently, delayed diagnosis or misdiagnosis may occur due to the lack of clarity in the diagnostic protocol (Geffner & Ross-Swain, 2018). However, early intervention for processing skills is significant as it contributes positively to school achievement (Cacace & McFarland, 1998; DeBonis & Moncrieff, 2008), language and speech development (Barrazo et al., 2016; Moore, 2007), cognitive abilities (Tomlin et al., 2015), and psychosocial status (Kreisman et al., 2012).

Therefore, it is important to use screening tools to identify individuals at risk, such as those with hearing loss, otitis media, communication disorders, neurodevelopmental disorders, neurological disorders, hereditary predispositions, or premature birth, and direct them to a comprehensive evaluation for auditory and speech processing skills (Bellis, 2011; Geffner & Ross-Swain, 2018). When the scanning tools are examined, there are tools by which recognition and processing skills can be evaluated in many ways, if not as a whole (Goldman, 2015; Reynolds et al., 2005). These screening tools include performance-based measurements and observations of families and teachers (Geffner & Ross-Swain, 2018; Musiek & Chermak, 2007). Additionally, processing time and nature can be determined using electrophysiological tests (ASHA, 2005a). However, performing these procedures in clinical settings does not take place because of limited time, equipment availability, and patience of the individual. Therefore, behavioural tests are preferred in clinical settings. It has been argued that auditory processing can be evaluated using tonal and speech stimuli as behavioural measures (Chermak et al., 1997; Masters et al., 1998). It is believed that tonal tests do not reflect daily life auditory processing skills, so it is recommended that auditory processing skills be evaluated using speech stimuli (Katz, 2016). However, auditory processing cannot be measured exactly due to the effect of cognitive loads when a speech stimulus is used (Musiek & Chermak, 2007).

In the literature, there are behavioural test batteries that screen and provide a comprehensive evaluation of auditory and speech processing skills. These test batteries include both tonal and speech stimuli. For example, the Test for Auditory Processing Disorders (SCAN; Keith, 2000) is an evaluation tool for adults and children with APD. It consists of four subtests: Filtered Words, Auditory Figure Ground, Competing Words, and Competing Sentences. A standardized assessment tool called the Language Processing Test-3 (Richard & Hanner, 2005) is used to evaluate various language processing skills in people between the ages of 5 and 21 and includes six subtests: Associations, Categorization, Similarities, Differences, Multiple Meanings, and Attributes. The Screening Test for Auditory Processing (STAP; Yathiraj & Maggu, 2013) was designed to scan children for APD. It includes Speech Perception in Noise, Dichotic Consonant-Vowel, Gap Detection, and Auditory Memory subtests.

The differences in responses to auditory stimuli in noisy and quiet backgrounds can reveal reaction times for processing (Houben et al., 2013; Meister et al., 2018; Rönnberg et al., 2013). To evaluate processing skills, individuals can respond to stimuli by speaking, pointing out, using eye movements, or pressing a button in the presence or absence of background noise. (Geffner & Ross-Swain, 2018; Katz, 2016; Keith, 2000; Martin & Brownell, 2005; Musiek & Chermak, 2007). Verbal responses and reaction times can be recorded for auditory or speech processing skills (Holden et al., 2019; Meister et al., 2018). Meister et al. (2018) investigated the verbal reaction time during the conventional speech-in-noise test, grouping the participants according to age and listening status, then examining noise types and intelligibility levels, and they stated that verbal reaction time could be easily evaluated during conventional speech audiometry.

In the literature, it is reported that the prevalence of APD in children ranges from 73% to 96% (Wilson & Arnott, 2013), with a male-to-female ratio of 2:1 (Chermak et al., 1997). In adults aged 55 and above, the prevalence varies from 23% to 76% (Golding et al., 2004). Considering the profound impact of auditory and speech processing skills on an individual's quality of life, a comprehensive assessment and early intervention are important when there is suspicion of a disorder in these skills (Geffner & Ross-Swain, 2018; Musiek & Chermak, 2007).

Therefore, there is a pressing need to develop an assessment tool that can rapidly screen individuals and is easily accessible to clinicians. This assessment tool would not only contribute to the positive outcomes of early intervention but also support comprehensive therapy programs. Furthermore, focusing on processing skills in more detail, especially in individuals with cognitive and sensory impairments, central auditory processing, and language and speech disorders, could contribute to the functional use of these dimensions, which received limited or no attention in therapy programs (Cacace & McFarland, 1998; Geffner & Ross-Swain, 2018; Katz, 2016)

A review of the literature reveals that screening tests, both observation-based and performance-based, are available for assessing auditory and speech processing. However, performance-based screening tests may have limitations in evaluating everyday life conditions and may require clinical settings for assessment. Consequently, there is a need for a screening test that evaluates processing based on real-life challenges. Furthermore, there is no experimental behavioural test to measure processing speed, even at the screening level in the literature. Because there are controversies in the evaluation of processing skills, it is important to develop performance screening tools that will indirectly reflect processing skills (Smoski et al., 1998). Considering auditory discrimination and recognition skills as performance tests (Archbold et al., 1998; Smoski et al., 1998), this study aimed to develop a screening method to determine auditory and speech skills.

As there is no gold standard for diagnosing APD, this study included individuals with various disorders that are thought to have similar findings as APD, instead of only those diagnosed with APD. Therefore, in addition to participants with typical development, individuals diagnosed with SLD, whose auditory processing abilities are known to be affected (Dawes & Bishop, 2010; Hämäläinen et al., 2013; King et al., 2003), also participated in this study. The literature indicates that individuals with typical development and SLD are commonly included as participants in APD studies. For instance, the development of the Comprehensive Test of Phonological Processing-2 (CTOPP-2; Wagner et al., 1999) involved 1900 participants aged 4 to 24 years with typical development, while the Auditory Skills Assessment (Geffner & Goldman, 2010) was conducted with 475 typical participants for test development. Domitz and Schow (2000) included individuals diagnosed with SLD as an atypical group during the development of the Multiple Auditory Processing Assessment-2 (MAPA-2), and Martin et al. (2018) included individuals diagnosed with SLD during the development of the Test of Auditory Processing-4.

The primary aim of this study was to develop a screening test called the Auditory and Speech Performance Test for Children (ASPT-C), which evaluates the auditory and speech performance of children 7.0 to 10;12 years old, with the parameters of recording correct/false and millisecond-rate reactions to meaningful and meaningless rhyming word pairs in the presence or absence of background noise. When psychoeducational tests developed to assess processing skills are examined, various validity parameters are evaluated in the validity study of these tests. For example, Webster (2009) assessed the content and construct validity of the Test of Information Processing Skills in a validity study, and Martin and Brownell (2005) examined the correlation between participants' test scores and their age and intelligence scores in a validity study of the Test of Auditory Processing-3. In this study, we conducted an assessment of the content, construct validity, and reliability of the ASPT-C.

Method

Participants

School-age children with SLD who complained of difficulty understanding in noise and children with typical development participated in this study. School-age children in the schools of the districts of Ankara, whose parents' consent was obtained and who participated voluntarily, were included in the study.

A total of 387 children participated in this methodological study: 307 children with typical development (Group 1) and 80 children with SLD (Group 2). The reason for the onset age of 7.0 was literacy and auditory processing maturity (Jerger & Musiek, 2000; Moore et al., 2011). This study was limited to elementary school children due to the challenges encountered in accessing schools as a result of the COVID-19 pandemic. Participants were selected using stratified sampling according to age, gender, and district.

Because the nonstandardized diagnostic tools of APD and the literature indicate that auditory processing is also affected in SLD (Alles et al., 2011; ASHA, 2005a; Bellis, 2011; Eggermont, 2015; Moore et al., 2011; Sharma et al., 2009), 80 school-age children with SLD who had complaints of difficulty understanding in noisy environments were included in the study by stratified sampling according to age and gender (Group 2). To determine the complaint of difficulty understanding noise, a data collection form prepared by the authors was used due to the lack of a valid and reliable scale for these ages in Türkiye. The data form included questions related to comprehension of speech, understanding of speech in noisy environments, following instructions, and history of language delay. The data form

was reviewed by five clinical audiologists working in research hospitals (3 females, age *M* = 27.3, *SD* = 2.8, professional experience *M* = 6.2 years, *SD* = 1.3) to ensure content validity.

The inclusion criteria for Group 1 were voluntary participants who were right-dominant handed, stated to have no disorders by their teachers or family, literate, and had consent from their families; Group 2 included voluntary participants who were right-dominant handed, literate, and had no diagnosis except for a diagnosis of SLD, complained of understanding in noise, and had consent from their families.

To assess the reading abilities of the groups, reading passages were read aloud to the children for 1 min, according to age-specific texts prepared by the authors and linguists. Participants who were at the age cutoff point were included in the study (for further details, see Erden et al., 2002).

Development of ASPT-C Items

The standards for reporting diagnostic accuracy (Bossuyt et al., 2015) guideline, in addition to reference sources (Boateng et al., 2018; de Vet et al., 2011; Streiner et al., 2015), were reviewed to guide both the study design and the development of the ASPT-C items.

After the literature review, the following four tasks were defined and developed for the ASPT-C. While developing the tasks of the ASPT-C, certain steps of listening skills were considered to reflect the recognition of speech in daily life (American Academy of Audiology [AAA], 2010; Archbold et al., 1998). These steps include detection, discrimination, recognition, and comprehension (Erber, 1975). The detection level was determined by the presence or absence of sounds. The discrimination level focused on similarities and differences in the sound (AAA, 2010; Estabrooks et al., 2016). Accordingly, it can be measured by showing or repeating what was heard between two similar sounds or by determining whether the minimal pairs were the same (Meinzen-Derr et al., 2007). The AAA (2010) recommended assessing the discrimination level of listening in the evaluation of auditory processing skills. In this context, the ability to distinguish was also examined using the ASPT-C. The recognition level is defined as the ability to perceive a stimulus under difficult listening conditions (Erber, 1975; Meinzen-Derr et al., 2007). Based on this definition, in the ASPT-C, meaningless items were created to simulate difficult listening conditions and to control for executive functions, working memory, and the listening effort effect in the evaluation of speech processing skills (Danneels et al., 2021; DeBonis, 2015).

In the evaluation of processing skills, it is recommended to evaluate the discrimination of speech in noise to reflect daily life (Archbold et al., 1998; Bellis, 2011; Yathiraj & Vanaja, 2018). When the quality of the signal decreases due to factors such as noise or hearing loss, speech processing can become more challenging, particularly when the language used is complex or the message content is less familiar (Akeroyd, 2008; Danneels et al., 2021; Libben et al., 2020). Therefore, noise stimuli were added to the speech-processing tasks in the ASPT-C, with noisy backgrounds that are often encountered in daily life, while presentation in quiet backgrounds was used for auditory performance assessment.

For the auditory discrimination subdimension, measures included the reaction times and the number of correct answers to meaningful monosyllabic rhyming words in silence; for the auditory recognition subdimension, the reaction times and the number of correct answers to meaningless monosyllabic rhyming words in silence; for the speech discrimination subdimension, the reaction times and the number of correct answers to meaningful monosyllabic rhyming words in background noise; and for the speech recognition subdimension, the reaction times and the number of correct answers to meaningless monosyllabic rhyming words in background noise.

Task items were prepared according to the manner, place, and voicing of phonemes and, as monosyllabic minimal pair words, changed sounds at the beginning and end of the word. When preparing meaningless monosyllabic word pairs, vowel-distributed words were chosen to prevent intelligibility skills and inferences from affecting the data (Steadman & Sumner, 2018). Prepared words were collected from a pool and checked by a linguist. Each task consisted of 15 items, and 60 items were developed. (see **Table 1**).

Audiovisual Materials

A male voice was used to record one of the words in each word pair in a sound studio. In this study, a male voice was used as the auditory stimulus following a comprehensive review of reference tests (Katz, 2016; Keith, 2000; Reynolds et al., 2005). The stimulus was presented at 50 dB HL with background noise added for speech tasks, and the signal-to-noise ratio was set to 0. This selection of 50 dB HL as the listening level was made because the most comfortable listening level would be at least 15 dB above background noise (Kobayashi et al., 2007; Ueda & Tanaka, 2020). In the ASPT-C, each word pair was presented twice, with and without background noise, and the answers were randomized. For instance, in

Table 1

Auditory and Speech Performance Test for Children Task Items

Number	Discrimination performance		Recognition performance	
	Auditory discrimination	Speech discrimination	Auditory recognition	Speech recognition
1	**bil**-pil	bil-**pil**	fim- **fom**	**fim**- fom
2	**var**-far	**var**-far	min-**mun**	min-**mun**
3	cam-**çam**	**cam**-çam	liş-**laş**	liş-**laş**
4	**tüh**-tüy	**tüh**-tüy	**yeç**-yıç	**yeç**-yıç
5	kıs-**kız**	kıs-**kız**	**sem**-sam	sem-**sam**
6	buz-**muz**	**buz**-muz	baf-**bif**	**baf**-bif
7	taç- **saç**	taç- **saç**	**çum**-çem	çum-**çem**
8	**et**-ek	et-**ek**	**pez**-pız	**pez**-pız
9	**bar**- bal	bar- **bal**	nef-**naf**	**nef**-naf
10	harf-**harp**	**harf**-harp	**fap**-fip	fap-**fip**
11	**çiz**-diz	çiz-**diz**	kuç-**keç**	**kuç**-keç
12	**pas**-tas	**pas**-tas	dız-**döz**	**dız**-döz
13	del-**gel**	del-**gel**	**ron**-rün	**ron**-rün
14	**hoş**- loş	**hoş**- loş	kım-**kem**	kım-**kem**
15	**puf**-pus	**puf**-pus	**şit**-şat	şit-**şat**

Note. Boldface values represent the correct answers.

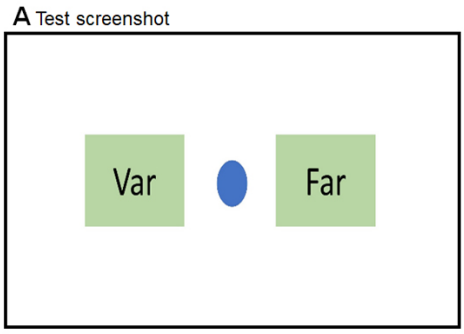
the word pair "fan-van," the correct answer was "fan" in the speech recognition task, while in the speech discrimination task, it could be "fan" or "van."

The type of background noise used in the study was "restaurant ambient." The reason for the use of this noise type was the reliability of the measurement by considering selective attention in the speech stimulus, which is sufficient to measure central auditory processing and is more convenient in daily life (Brungart & Simpson, 2007; Evans et al., 2016; Holube, 2011). The background noise (Stephan, n.d.) was obtained from the soundbible.com website where free audio files are available. The necessary arrangements were made as mentioned in the literature (Leccese et al., 2015; Zokoll et al., 2015) before superimposing the noise on the male voice.

Sound recordings were performed in a sound studio with the help of a sound technician. The devices used for recording and editing were Sound Recorder (Shure PG42-LC), a condenser microphone and speaker (Sony), and Sound Forge PRO 10 (Sony). The soundboard was a Soundcraft FX16. During sound recording, the microphone distance from the speaker was maintained at a minimum of 30 cm by positioning the microphone for optimal directionality; mono recordings were digitized at a sampling rate of 44.1 kHz with 16-bit amplitude resolution. The Audition 11.0 (Adobe) program was used to combine the stimuli prepared in the studio with the noise stimulus. The program used in arranging the noise characteristics and preventing sound explosions was based on Leccese et al.'s (2015) study, which utilized a pop-up filter.

The test items were prepared using written visuals. In the test, hand-eye coordination was not considered, and the word pairs on the screen were positioned perceptually on the right and left simultaneously in the middle of the screen so that the visual sense did not affect the results of the test (Jain et al., 2015). Participants were asked to perform this procedure by looking at the blue dot between two written words (Amini Vishteh et al., 2019). Participants could change the displayed word pairs by pressing a predefined button (see **Figure 1**).

Figure 1



B Test result output

AUDIO TEST Speech+Cog				
PROTCL:				
ID:				
AGE:				
SEX:				
HANDEDNESS:		Qs # True	Qs # False	Qs Interval
# EVENT:	60.00	8	7	2885.680
# TRUE:	31.00	7	8	3054.030
# FALSE:	29.00	9	6	2431.595
# FAULT:	0.00	7	8	2594.655
T. TIME, ms:	164489.40	A=	-168.35	
AVER, ms:	2741.49	B=	-163.06	

Screenshots of Auditory and Speech Performance Test for Children (ASPT-C) and data output

Note. Panel A: Screenshot of ASPT-C. Minimal pairs were located on the right and left sides of the midline of the screen. Participants listened to the stimulus while looking at the blue dot and pressed the keyboard key assigned to the appropriate written stimulus for the auditory stimulus. Panel B: Data output. The output included scores for data (reaction times and accuracy on subtasks) and participant characteristics.

Software

Auditory and visual stimuli were embedded into the digital finger-tapping test battery (Kiziltan et al., 2006), which measures tapping performance with a high time resolution and saves the data on the hard drive of a computer for analysis. The software was developed for use on a standard personal computer without the need for auxiliary hardware. The system used a read-time stamp counter, which is a powerful benchmark introduced by Intel in Pentium processors (Intel, Santa Clara, CA). Therefore, in an IBM-compatible personal computer with at least 1 GHz CPU, it is possible to reach a time resolution of milliseconds in measuring finger-tapping tasks (Aydin et al., 2016). With the new version of the software, it became possible to measure reaction times and responses to audiovisual stimuli. Participants were asked to press the predefined key on the computer keyboard as soon as they answered the audiovisual task. The test module automatically calculated the average reaction times and correct answers for the tasks and total test (see **Figure 1**).

Procedure

This study was conducted with ethical approval from the Ankara University Directorate (17.12.2021, I11-693-21). The study followed the principles of the Declaration of Helsinki with specific measures implemented to ensure participant confidentiality, informed consent, and fair treatment. Data collection was authorized by the Ankara Governorship Directorate of National Education (20.11.2020, 14588481-605.99-E.17020975).

All tests were conducted on a Lenovo Desktop-UTOJTA5 notebook with an Intel Core i5-4210U CPU @ 1.70 GHz 2.40 GHz processor and 64-bit operating system. The tests were administered in quiet locations within schools, such as the library or manager's office.

Before the main procedure, a preliminary study was conducted with 20 children with typical development (10 females; age $M = 8.52$, $SD = 2.1$) to ensure that the instructions, audiovisual materials, and test items were appropriate.

During the main procedure, each participant sat comfortably in front of the computer screen and was instructed to press a predefined keyboard key in response to each task. This instruction was presented to each participant.

Statistical Analyses

Regarding the psychometric properties of the test, answers to the research questions about the validity and reliability of the test were sought.

In the validity study, the content validity ratio (CVR) and content validity index (CVI) were determined for content validity. A correlation matrix suitable for exploratory factor analysis (EFA), Kaiser-Meyer-Olkin (KMO), Bartlett's sphericity test, and common variance values were determined and EFA was performed. As the data did not have a normal distribution for discriminant validity, differences between groups were calculated using the Mann-Whitney U test and multigroup differences were calculated using the Kruskal-Wallis test. As there was a

difference, the Mann-Whitney U test (with the corrected Bonferroni test) was used to determine the groups from which the difference originated. Student's *t* test was used for item discrimination.

For internal consistency, Cronbach's alpha and the Kuder Richardson-20 (KR-20) reliability coefficient were used, and for test-retest reliability, the intraclass correlation coefficient with the two-way mixed model was analyzed.

Results

Participants

Of the participants, 39.8% (*n* = 154) were female and 60.2% (*n* = 233) were male. When analyzed by age, 17.05% (*n* = 66) were between 7.0 and 7;12 years old, 30.24% (*n* = 117) were between 8.0 and 8;12 years old, 29.20% (*n* = 113) were between 9.0 and 9;12 years old, and 23.51% (*n* = 91) were between 10.0 and 10;12 years old. According to socioeconomic status, 33.3% (*n* = 129) were classified as low, 34.7% (*n* = 134) as moderate, and 32% (*n* = 124) as high. According to grade level, 16.80% (*n* = 65) were in the 2nd grade, 29.71% (*n* = 115) in the 3rd grade, 29.97% (*n* = 116) in the 4th grade, and 23.52% (*n* = 91) in the 5th grade (see **Table 2**).

Content Validity

Expert opinion was sought to determine the content validity of the tests recorded by the CVR and CVI. Accordingly, the CVR for the test items were calculated with the formulation [G/(N/2)]–1 (where G = number of experts scoring, 4+5/3+4+5 and *N* = total number of experts). CVI is the average CVR value of the items remaining in the pool (Lawshe, 1975). For expert opinions, the CVR of test items were 1.0 for the Auditory Discrimination task; .94 for the Speech Discrimination task; .89 for the Auditory Recognition task; .92 for the Speech Recognition task. CVI values were found 1.0 in the Auditory Discrimination task; .94 in the Speech Discrimination task; .89 in the Auditory Recognition task; and .92 in the Speech Recognition task.

Extraction of Factors

EFA was employed to assess the factors that fit the items. (Boateng et al., 2018; DeVellis & Thorpe, 2021). Before applying EFA, the suitability of the items for the factor structures of the test was checked using a correlation matrix. According to the test tasks, the correlation matrix of the test items had a factor load greater than .30.

Table 2

Participant Characteristics

Characteristics	Group 1		Group 2		Total	
	n	%	*n*	%	*n*	%
Gender						
Female	123	40.00	31	38.75	154	39.80
Male	184	60.00	49	61.25	233	60.20
Ages in years; months						
7;0–7;12	53	17.26	13	16.25	66	17.05
8;0–8;12	92	29.96	25	31.25	117	30.24
9;0–9;12	90	29.32	23	28.75	113	29.20
10;0–10;12	72	23.46	19	23.75	91	23.51
Socioeconomic status						
Low	103	33.60	26	32.50	129	33.30
Middle	105	34.20	29	36.25	134	34.70
High	99	32.20	25	31.25	124	32.00
Grade						
2	52	16.94	13	16.25	65	16.80
3	90	29.32	25	31.25	115	29.71
4	95	30.94	21	26.25	116	29.97
5	70	22.80	21	26.25	91	23.52

The suitability of the ASPT-C data for factor analysis was examined using the KMO coefficient and Bartlett's sphericity test. In this study, the KMO value was .79, and Bartlett's Sphericity test yielded significant results ($x^2(6) = 717.594$, $p < .001$). After determining whether the test was suitable for factor analysis, the common variances of the test items were examined and the values were found to be sufficient for the factor analysis. ASPT-C gathered two factors with an eigenvalue of 1.92, accounting for 66.65% of the total variance. The first factor was named Accuracy and the second factor was named Reaction Time (see **Table 3**).

Item Discrimination

To evaluate the item discrimination of the ASPT-C, the difference between the scores in the top 27% and bottom 27% was examined. For the Accuracy subtest, a significant difference was found between the bottom 27% ($n = 104$, $M = 50.24$, $SD = 5.91$) and top 27% ($n = 104$, $M = 58.44$, $SD = 0.57$), $t(207) = -14.13$, $p < .001$, $r = 1.95$. When the Reaction Time data were examined, a significant difference was found between the bottom 27% ($n = 104$, $M = 1892.33$, $SD = 110.69$) and the top 27% ($n = 104$, $M = 2667.29$, $SD = 258.91$), $t(207) = -28.09$, $p < .001$, $r = 3.89$; see **Table 3** for more detail).

Construct Validity

Construct validity was assessed through differentiation based on "known groups" by analyzing the differences among groups, ages, and task scores.

To compare the Accuracy and Reaction Time scores between Group 1 ($M = 56$, $SD = 3.8$ for Accuracy, $M = 2175.1$, $SD = 285.8$ for Reaction Time) and Group 2 ($M = 51.4$, $SD = 4.9$ for Accuracy, $M = 2483$, $SD = 375.5$ for Reaction Time), both the total test and tasks were considered (see **Table 4**). The results showed that the Accuracy and Reaction Time scores were statistically significant ($U = 3894.5$, $z = -9.48$, $p < .001$, $r = .48$ for Accuracy total score; $U = 6339$, $z = -6.66$, $p < .001$, $r = .33$ for Reaction Time total score).

When the background noise in the discrimination subtasks of Group 1 was examined, the difference between the presence of background noise ($M = 13.5$, $SD = 1.2$) and

Table 3

Total Variance Results and Eigenvalue Coefficient of Auditory and Speech Performance Test for Children

Tasks	Factor loading		Top 27% and bottom 27% group comparison
	1	2	*t*
Factor 1: Accuracy, Eigenvalue = 4.74, Total variance = 37.8%			
Auditory Discrimination	−.148	**.686**	−16.25* (df=208)
Speech Discrimination	−.236	**.611**	−21.23* (df=208)
Auditory Recognition	−.068	**.782**	−15.37* (df=208)
Speech Recognition	−.153	**.689**	−27.77* (df=208)
Factor 1's total score	**−.184**	**.887**	**−14.13* (df=207)**
Factor 2: Reaction Time, Eigenvalue = 1.92, Total variance = 28.85%			
Auditory Discrimination	**.801**	−.122	−24.22* (df=207)
Speech Discrimination	**.809**	−.210	−24.67* (df=207)
Auditory Recognition	**.820**	−.163	−19.40* (df=207)
Speech Recognition	**.849**	−.198	−23.27* (df=207)
Factor 2's total score	**.975**	**−.208**	**−28.09* (df=207)**
Total Eigenvalue = 1.92, Total variance = 66.65%			

Note. Factor analysis after varimax rotation. Boldface values represent primary loading associated with each factor.
*$p < .001$

Table 4

Comparison Between Groups According to Accuracy and Reaction Time

Task	Subtask	Group 1		Group 2		*U*	*z*	*r*
		M ± *SD*	*z*-scores	*M* ± *SD*	*z*-scores			
Accuracy	Auditory Disc.	14.4 ± 0.9	−9.81*	13.3 ± 1.9	−5.48*	7273.5	−6.21*	.31
	Speech Disc.	13.5 ± 1.2		11.7 ± 1.7		4574.5	−8.86*	.53
	Auditory Recog.	14.6 ± 0.7	−10.58*	13.8 ± 1.5	−5.86*	8332.5	−5.22*	.26
	Speech Recog.	13.8 ± 1.0		12.7 ± 1.8		7355.5	−5.74*	.29
	Total score	56 ± 3.8		51.4 ± 4.9		3894.5	−9.48*	.48
Reaction time (ms)	Auditory Disc.	2169.3 ± 330.6	−10.65*	2393.9 ± 478.9	−6.20*	8715	−4.00*	.20
	Speech Disc.	2379.3 ± 385.8		2759.9 ± 472.4		6325	−6.68*	.33
	Auditory Recog.	1935.2 ± 316.5	−13.23*	2277.1 ± 488.5	−5.25*	6910	−6.02*	.30
	Speech Recog.	2206.2 ± 330.7		2500.9 ± 441.8		7271	−5.62*	.28
	Total score	2175.1 ± 285.8		2483 ± 375.5		6339	−6.66*	.33

Note. Disc = Discrimination; Recog = Recognition. z-scores are based on the presence of noise stimulus in subtests.
* $p < .001$

absence of background noise ($M = 14.4$, $SD = 0.9$) in the Accuracy subtest, and the presence of background noise ($M = 2379.3$, $SD = 385.8$) and the absence of background noise ($M = 2169.3$, $SD = 330.6$) in the Reaction Time subtests were significant ($z = -9.81$, $p < .001$, $r = .56$ for Accuracy; $z = -10.65$, $p < .001$, $r = .6$ for Reaction Time). When considered as recognition subtasks, the scores between the presence of background noise ($M = 13.8$, $SD = 1.0$) and absence of background noise ($M = 14.6$, $SD = 0.7$) in the Accuracy subtest and between the presence of background noise ($M = 2206.2$, $SD = 330.7$), and the absence of background noise ($M = 1935.2$, $SD = 316.5$) in the Reaction Time subtest was also significant ($z = -10.58$, $p < .001$, $r = .6$ for Accuracy; $z = -13.23$, $p < .001$, $r = .75$ for Reaction Time). When the background noise in the discrimination subtasks of group 2 was examined, the difference between the presence of background noise ($M = 11.7$, $SD = 1.7$) and absence of background noise ($M = 13.3$, $SD = 1.9$) in the Accuracy subtest, and presence of background noise ($M = 2759.9$, $SD = 472.4$) and absence of background noise ($M = 2393.9$, $SD = 478.9$) in the Reaction Time subtest were significant ($z = -5.48$, $p < .001$, $r = .6$ for Accuracy; $z = -6.20$, $p < .001$, $r = .69$ for Reaction Time). Likewise, in the recognition subtest, the scores between the presence of background noise ($M = 12.7$, $SD = 1.8$) and absence of background noise ($M = 13.8$, $SD = 1.5$) in the Accuracy subtest, and between the presence of background noise ($M = 2500.9$, $SD = 441.8$) and the absence of background noise ($M = 2277$, $SD = 488.4$) in Reaction Time were found to be significant ($z = -5.86$, $p < .001$, $r = .65$ for Accuracy; $z = -5.25$, $p < .001$, $r = .58$ for Reaction Time; see **Table 4**). Accordingly, when the noise stimulus was added, the accuracy decreased and the reaction time was delayed.

When analyzing the ASPT-C scores according to age, significant differences were found in the Accuracy scores of Group 1 ($H(3) = 49.20$, $p < .001$). Pairwise comparisons showed differences between the ages of 7 and 9 ($z = -52.18$, $p < .001$, $r = 4.3$), between the ages of 7 and 10 ($z = -88.09$, $p < .001$, $r = 8$), between the ages of 8 and 9 ($z = -44.26$, $p < .001$), and between the ages of 8 and 10 ($z = -80.16$, $p < .001$, $r = 6.2$). Significant differences were also found in the Reaction Time scores between ages ($H(3) = 110.64$, $p < .001$). Pairwise comparisons showed differences between the ages of 7 and 9 ($z = 68.19$, $p < .001$, $r = 5.6$), between the ages of 7 and 10 ($z = 114.28$, $p < .001$), between the ages of 8 and 9 ($z = 40.75$, $p < .001$, $r = 3.7$), between the ages of 8 and 10 ($z = 116.84$, $p < .001$, $r = 9$), and between the ages of 9 and 10 years ($z = 76.09$, $p < .001$, $r = 6$). In Group 2, there were no significant differences in Accuracy and Reaction Time scores between ages ($H(3) = 8.287$, $p = .040$ for Accuracy; $H(3) = 4.931$, $p = .177$ for Reaction Time).

In summary, ASPT-C has content validity. It has a two-factor structure: Reaction Time and Accuracy. The ASPT-C is distinctive according to groups, age, and subtest.

Reliability

Cronbach's alpha and KR-20 were used to assess the internal consistency and reliability of the ASPT-C test. The Cronbach's alpha values of the ASPT-C's Reaction Time tasks ranged from .77 to .90, indicating moderate to high internal consistency (see **Table 5**). The total scale had a

Table 5

Auditory and Speech Performance Test for Children Cronbach's Alpha (α) Values for Internal Consistency

Tasks	*M* ± *SD*	α (*n* = 387)	ICC (*n* = 120)	Confidence interval
Auditory Discrimination	2215.69 ± 376.64	.77	.67	.47–.79
Speech Discrimination	2458.03 ± 433.03	.78	.76	.62–.84
Auditory Recognition	2005.86 ± 384.03	.90	.80	.68–.87
Auditory Recognition	2267.14 ± 375.37	.80	.74	.58–.84
Total score	2238.75 ± 330.47	.93	.93	.90–.96

Note. ICC = intraclass correlation coefficient

Cronbach's alpha of .93, indicating high reliability. Similarly, the KR-20 coefficient for the Accuracy tasks was .70, suggesting moderate reliability (see **Table 6**). However, the KR-20 values for the subscores ranged from .20 to .56, indicating low internal consistency.

Test-retest reliability was assessed using the intraclass correlation coefficient, which ranged from .67 to .80 for Reaction Time and .49 to .76 for Accuracy. The correlation coefficients were .93 for Reaction Time ($F(79) = 11.36$, $p < .001$) and .89 for Accuracy ($F(79) = 16.06$, $p < .001$), indicating high reliability for the total scores.

In summary, the subtasks of Accuracy and Reaction Time were moderately reliable, and the total scores of Accuracy and Reaction Time were highly reliable.

Discussion

In this study, 307 school-age children with typical development and 80 school-age children with SLD were evaluated for auditory and speech performance using the ASPT-C. The test measured correct answer scores and reaction times in milliseconds in response to meaningful and meaningless minimal pairs, in both the presence and absence of background noise. Content validity, factor extraction, item discrimination, construct validity, internal consistency, and test-retest reliability of the ASPT-C were analyzed.

When developing a scale or test, it is important to determine validity and reliability. EFA, one of the methods used to identify factor loads before establishing the construct validity of the assessment tool, was conducted (DeVellis & Thorpe, 2021). Upon reviewing the literature, EFA was used for the construct validity of the STAP (Yathiraj & Maggu, 2013), which had three factors, and Domitz and Schow (2000) revealed that MAPA had four factors. In this study, we developed a test to explain these two factors by performing EFA. The ASPT-C consisted of four subtasks, each with two response categories (accuracy and reaction times). These response categories accounted for two-factor loading, and demonstrated the construct validity of the ASPT-C.

To enhance the construct validity of the ASPT-C, participant scores in the lower and upper 27% groups were compared, revealing item discrimination. Item discrimination is used to strengthen the construct validity of the items in scales and/or tests (Johnston et al., 2014). When reviewing the literature related to the assessment of auditory or speech processing skills, no scale/test was found that demonstrated item discrimination. In this regard, we introduced a test with established item discrimination values in the literature.

A comparison of the test with an atypical group is also an important parameter for construct validity. Schow et al. (2021) included participants diagnosed with speech difficulties, attention-deficit/hyperactivity disorder, and dyslexia in the MAPA-2 test, as well as individuals with normal development in their study. Keith (2000) included individuals diagnosed with APD as both a typical developmental and atypical group in a validity test for children with APD (SCAN 3-C). In this study, we included children with SLD who had complaints of difficulty understanding in noisy environments. By comparing the two groups, statistically significant differences were found in test scores between children with typical development and those with SLD. This result demonstrated that ASPT-C reveals the difference between the groups.

To demonstrate the construct validity, age groups, which are thought to affect auditory and speech processing skills, were compared. When these findings were examined, differences were observed among the age groups. It could be seen that as age increased, both the reaction speed and the number of correct responses increased. This finding

Table 6

Auditory and Speech Performance Test for Children KR-20 Values for Internal Consistency

Tasks	*M* ± *SD*	KR-20 coefficient (*n* = 387)	ICC (*n* = 120)	Confidence interval
Auditory Discrimination	14.18 ± 1.26	.56	.56	.32–.72
Speech Discrimination	13.12 ± 1.53	.39	.76	.63–.84
Auditory Recognition	14.43 ± 0.98	.44	.68	.50–.79
Auditory Recognition	13.53 ± 1.27	.20	.49	.20–.68
Total score	55.11 ± 4.44	.70	.89	.79–.94

Note. ICC = intraclass correlation coefficient

confirms the hypothesis that processing skills improve with age (Jerger & Musiek, 2000; Moore et al., 2011).

Children with auditory and speech processing disorders may encounter difficulties in understanding instructions in noisy environments in their daily lives. Therefore, noise stimuli are frequently employed to assess APD. In this study, we structured the ASPT-C subtasks based on the presence or absence of noisy stimuli. We found an increase in reaction times and incorrect responses with background noise, both among children with typical development and those with SLD. When comparing the groups, we found that children with SLD struggled more with these tasks. Ferenczy et al. (2022), Koiek et al. (2018), and Warrier et al. (2004) found significant differences in comprehension scores of individuals with SLD under noisy conditions. The results of this study demonstrate the effectiveness of noise stimuli in measuring processing skills and contribute to the extension of findings in the literature.

Internal consistency and intraclass correlation coefficients were calculated to examine the reliability of the ASPT-C. In the literature, reliability analyses of APD tests show highly reliable results for certain tests and tasks. However, some of these tests may have low reliabilities. In the reliability study of the Feather Squadron Test (Barker & Purdy, 2016), scores varied between 17.3 and 90.8 depending on the tasks. The reliability of CTOPP-2 (Wagner et al., 1999) tasks was found to have test-retest correlations between .75 and .92 for core tasks, between .76 and .86 for composites, and .73 and .75 for tasks. These low or medium reliability results were attributed to factors such as a lack of maturation in auditory processing, difficulty in directing attention, and listening effort. The reliability of the ASPT-C was also examined using internal consistency and stability parameters and was found to be moderately or highly reliable. This result is likely to have affected the reliability of the 7-year-old age group. In addition, we believe that the test was performed on a computer and that the attention of the participants, listening effort, and age were important factors.

The ASPT-C has demonstrated its potential for assessing auditory and speech processing skills in children. Consequently, it is believed that this test can be employed to screen for auditory and speech processing disorders and facilitate early intervention. ASPT-C offers practicality and ease of use as a noteworthy advantage. Additionally, this study is anticipated to contribute to further research in the field of auditory and speech processing skills. Although our findings provide insights into the impact of auditory processing skills on individuals with SLD, further research is necessary to generalize these results.

The limitations of this study are that the study sample was limited to Türkiye, and the ASPT-C was performed with a single computer to ensure proper calibration of the test.

Finally, we conclude that the ASPT-C is a valid and reliable screening test that can be applied to children aged 7.0–10.12 years. We recommend performing the test with a larger sample size, comparing the results with those of different groups of people with varying needs and ages, and using a device specifically designed and calibrated for the test to avoid calibration problems. Additionally, it is recommended to compare the scores obtained from different processing tests with those obtained from the ASPT-C.

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47*(sup2), S53–S71. https://doi.org/10.1080/14992020802301142

Alles, R., Bamiou, D., Batchelor, L., Campbell, N. G., Canning, D., Grant, P., Luxon, L., Moore, D., Murray, P., & Nairn, S. (2011, March 31). *British Society of Audiology position statement: Auditory processing disorder (APD).* https://eprints.soton.ac.uk/338076/1/BSA_APD_PositionPaper_31March11_FINAL.pdf

American Academy of Audiology. (2010, August). *Diagnosis, treatment and management of children and adults with central auditory processing disorder* [Clinical practice guidelines]. https://audiology-web.s3.amazonaws.com/migrated/CAPD%20Guidelines%208-2010.pdf_539952af956c79.73897613.pdf

American Speech-Language-Hearing Association. (2005a). *Central auditory processing disorder*. https://www.asha.org/Practice-Portal/Clinical-Topics/Central-Auditory-Processing-Disorder/

American Speech-Language-Hearing Association. (2005b). *(Central) auditory processing disorders—The role of the audiologist* [Position statement]. https://doi.org/10.1044/policy.PS2005-00114

Amini Vishteh, R., Mirzajani, A., Jafarzadehpour, E., & Darvishpour, S. (2019). Evaluation of simple visual reaction time of different colored light stimuli in visually normal students. *Clinical Optometry, 11*, 167–171. https://doi.org/10.2147%2FOPTO.S236328

Archbold, S., Lutman, M. E., & Nikolopoulos, T. (1998). Categories of auditory performance: Inter-user reliability. *British Journal of Audiology, 32*(1), 7–12. https://doi.org/10.3109/03005364000000045

Aydin, L., Kiziltan, E., & Gundogan, N. U. (2016). Polyphasic temporal behavior of finger-tapping performance: A measure of motor skills and fatigue. *Journal of Motor Behavior, 48*(1), 72–78. https://doi.org/10.1080/00222895.2015.1040111

Barker, M. D., & Purdy, S. C. (2016). An initial investigation into the validity of a computer-based auditory processing assessment (Feather squadron). *International Journal of Audiology, 55*(3), 173–183. https://doi.org/10.3109/14992027.2015.1074734

Barrozo, T. F., Pagan-Neves, L. D. O., Vilela, N., Carvallo, R. M. M., & Wertzner, H. F. (2016). The influence of (central) auditory processing disorder in speech sound disorders. *Brazilian Journal of Otorhinolaryngology, 82*(1), 56–64. https://doi.org/10.1016/j.bjorl.2015.01.008

Bellis, T. J. (2011). *Assessment and management of central auditory processing disorders in the educational setting: From science to practice*. Plural Publishing.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, Article 149. https://doi.org/10.3389/fpubh.2018.00149

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., Lijmer, J.G., Moher, D., Rennie, D., de Vet, H. C. W., Kressel, H. Y., Rifai, N., Golub, R. M., Altman, D. G., Hooft, L., Korevaar, D. A., & Cohen, J. F. (2015). STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ, 351*, Article 5527. https://doi.org/10.1136/bmj.h5527

Brungart, D. S., & Simpson, B. D. (2007). Cocktail party listening in a dynamic multitalker environment. *Perception & Psychophysics, 69*(1), 79–91. https://doi.org/10.3758/BF03194455

Cacace, A. T., & McFarland, D. J. (1998). Central auditory processing disorder in school-aged children: A critical review. *Journal of Speech, Language, and Hearing Research, 41*(2), 355–373. https://doi.org/10.1044/jslhr.4102.355

Chermak, G. D., Musiek, F. E., & Craig, C. H. (1997). *Central auditory processing disorders: New perspectives*. Singular.

Danneels, M., Degeest, S., Dhooge, I., & Keppler, H. (2021). Central auditory processing and listening effort in normal-hearing children: A pilot study. *International Journal of Audiology, 60*(10), 739–746. https://doi.org/10.1080/14992027.2021.1877365

Dawes, P., & Bishop, D. V. (2010). Psychometric profile of children with auditory processing disorder and children with dyslexia. *Archives of Disease in Childhood, 95*(6), 432–436. https://doi.org/10.1136/adc.2009.170118

DeBonis, D. A. (2015). It is time to rethink central auditory processing disorder protocols for school-aged children. *American Journal of Audiology, 24*(2), 124–136. https://doi.org/10.1044/2015_aja-14-0037

DeBonis, D. A., & Moncrieff, D. (2008). Auditory processing disorders: An update for speech-language pathologists. *American Journal of Speech-Language Pathology, 17*(1), 4–18. https://doi.org/10.1044/1058-0360(2008/002)

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage Publications.

de Vet, H. C., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide*. Cambridge University Press. https://doi.org/10.1017/CBO9780511996214

Domitz, D. M., & Schow, R. L. (2000). A new CAPD battery—Multiple auditory processing assessment. *American Journal of Audiology, 9*(2), 101–111. https://doi.org/10.1044/1059-0889(2000/012)

Eggermont, J. J. (2015). *Auditory temporal processing and its disorders*. OUP Oxford.

Erber, N. P. (1975). Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders, 40*(4), 481–492. https://doi.org/10.1044/jshd.4004.481

Erden, G., Kurdoğlu, F., & Uslu, R. (2002). Ilköğretim okullarina devam eden Türk çocuklarinin sinif düzeylerine göre okuma hizi ve yazim hatalari normlarinin geliştirilmesi [Development of reading speed and spelling error norms according to the grade levels of Turkish children attending primary schools]. *Türk Psikiyatri Dergisi, 13*(1), 5–13.

Estabrooks, W., MacIver-Lux, K., & Rhoades, E. A. (2016). *Auditory-verbal therapy: For young children with hearing loss and their families, and the practitioners who guide them*. Plural Publishing.

Evans, S., McGettigan, C., Agnew, Z. K., Rosen, S., & Scott, S. K. (2016). Getting the cocktail party started: Masking effects in speech perception. *Journal of Cognitive Neuroscience, 28*(3), 483–500. https://doi.org/10.1162/jocn_a_00913

Ferenczy, M., Pottas, L., & Soer, M. (2022). Speech perception in noise in children with learning difficulties: A scoping review. *International Journal of Pediatric Otorhinolaryngology, 156*, Article 111101. https://doi.org/10.1016/j.ijporl.2022.111101

Geffner, D., & Goldman, R. (2010). *ASA: Auditory Skills Assessment*. PsychCorp.

Geffner, D., & Ross-Swain, D. (2018). *Auditory processing disorders: Assessment, management, and treatment*. Plural Publishing.

Golding, M., Carter, N., Mitchell, P., & Hood, L. (2004). Prevalence of central auditory processing (CAP) abnormality in an older Australian population: The Blue Mountains hearing study. *Journal of the American Academy of Audiology, 15*(9), 633–642. https://doi.org/10.3766/jaaa.15.9.4

Goldman, R. (2015). *Goldman-Fristoe Test of Articulation–Third edition (GFTA-3)*. AGS.

Hämäläinen, J. A., Salminen, H. K., & Leppänen, P. H. (2013). Basic auditory processing deficits in dyslexia: Systematic review of the behavioral and event-related potential/field evidence. *Journal of Learning Disabilities, 46*(5), 413–427. https://doi.org/10.1177/0022219411436213

Holden, J., Francisco, E., Lensch, R., Tommerdahl, A., Kirsch, B., Zai, L., Dennis, R., & Tommerdahl, M. (2019). Accuracy of different modalities of reaction time testing: Implications for online cognitive assessment tools. *BioRxIV*, Article 726364. https://doi.org/10.1101/726364

Holube, I. (2011). Speech intelligibility in fluctuating maskers. *Proceedings of the International Symposium on Auditory and Audiological Research, 3*, 57–64. https://proceedings.isaar.eu/index.php/isaarproc/article/view/2011-06

Houben, R., van Doorn-Bierman, M., & Dreschler, W. A. (2013). Using response time to speech as a measure for listening effort. *International Journal of Audiology, 52*(11), 753–761. https://doi.org/10.3109/14992027.2013.832415

Jain, A., Bansal, R., Kumar, A., & Singh, K. (2015). A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *International Journal of Applied & Basic Medical Research, 5*(2), 124–127. https://doi.org/10.4103/2229-516x.157168

Jerger, J., & Musiek, F. (2000). Report of the consensus conference on the diagnosis of auditory processing disorders in school-aged children. *Journal of the American Academy of Audiology, 11*(09), 467–474. https://doi.org/10.1055/s-0042-1748136

Johnston, M., Dixon, D., Hart, J., Glidewell, L., Schröder, C., & Pollard, B. (2014). Discriminant content validity: A quantitative methodology for assessing content of theory-based measures, with illustrative applications. *British Journal of Health Psychology, 19*(2), 240–257. https://doi.org/10.1111/bjhp.12095

Katz, J. (2016). The Buffalo CAPD Model: The importance of phonemes in evaluation and remediation. *Journal of Phonetics & Audiology, 2*(1), Article 111. https://doi.org/10.4172/2471-9455.1000111

Keith, R. W. (2000). Development and standardization of SCAN-C test for auditory processing disorders in children. *Journal of the American Academy of Audiology, 11*(08), 438–445. https://doi.org/10.1055/s-0042-1748131

King, W. M., Lombardino, L. J., Crandell, C. C., & Leonard, C. M. (2003). Comorbid auditory processing disorder in developmental dyslexia. *Ear and Hearing, 24*(5), 448–456. https://doi.org/10.1097/01.aud.0000090437.10978.1a

Kiziltan, E., Barut, C., & Gelir, E. (2006). A high-precision, low cost system for evaluating finger-tapping tasks. *International Journal of Neuroscience, 116*(12), 1471–1480. https://doi.org/10.1080/00207450500514318

Kobayashi, M., Morimoto, M., Sato, H., & Sato, H. (2007). Optimum speech level to minimize listening difficulty in public spaces. *The Journal of the Acoustical Society of America, 121*(1), 251–256. https://doi.org/10.1121/1.2382499

Koiek, S., Pourbakht, A., Mahdavi, M. E., & Tahaei, A. A. (2018). Acceptable noise level in learning disordered children. *Auditory and Vestibular Research, 27*(2), 86–92. https://avr.tums.ac.ir/index.php/avr/article/view/689

Kreisman, N., John, A., Kreisman, B., Hall, J., & Crandell, C. (2012). Psychosocial status of children with auditory processing disorder. *Journal of the American Academy of Audiology, 23*(3), 222–233. https://doi.org/10.3766/jaaa.23.3.8

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Leccese, F., Tuoni, G., Salvadori, G., & Rocca, M. (2015). An analytical model to evaluate the cocktail party effect in restaurant dining rooms: A case study. *Applied Acoustics, 100*, 87–94. https://doi.org/10.1016/j.apacoust.2015.06.012

Libben, G., Gagné, C. L., & Dressler, W. U. (2020). The representation and processing of compounds words. In V. Pirrelli, I. Plag, & W. U. Dressler (Eds.), *Word knowledge and word usage* (pp. 336–352). De Gruyter Mouton. https://doi.org/10.1515/9783110440577-009

Martin, N., & Brownell, R. (2005). *TAP-3: A language processing skills assessment.* WPS Publishing.

Martin, N., Brownell, R., & Hamaguchi, P. (2018). *TAP-4: A language processing skills assessment.* WPS Publishing.

Masters, M. G., Stecker, N. A., & Katz, J. (1998). *Central auditory processing disorders: Mostly management.* Pearson.

Meinzen-Derr, J., Wiley, S., Creighton, J., & Choo, D. (2007). Auditory skills checklist: Clinical tool for monitoring functional auditory skill development in young children with cochlear implants. *Annals of Otology, Rhinology & Laryngology, 116*(11), 812–818. https://doi.org/10.1177/000348940711601104

Meister, H., Rählmann, S., Lemke, U., & Besser, J. (2018). Verbal response times as a potential indicator of cognitive load during conventional speech audiometry with matrix sentences. *Trends in Hearing, 22*, Article 233121651879325. https://doi.org/10.1177/2331216518793255

Moore, D. R. (2007). Auditory processing disorders: Acquisition and treatment. *Journal of Communication Disorders, 40*(4), 295–304. https://doi.org/10.1016/j.jcomdis.2007.03.005

Moore, D. R., Cowan, J. A., Riley, A., Edmondson-Jones, A. M., & Ferguson, M. A. (2011). Development of auditory processing in 6- to 11-yr-old children. *Ear and Hearing, 32*(3), 269–285. https://doi.org/10.1097/AUD.0b013e318201c468

Musiek, F. E., & Chermak, G. D. (2007). *Handbook of central auditory processing disorder.* Plural Publishing.

Palana, J., Schwartz, S., & Tager-Flusberg, H. (2022). Evaluating the use of cortical entrainment to measure atypical speech processing: A systematic review. *Neuroscience & Biobehavioral Reviews, 133,* Article 104506. https://doi.org/10.1016/j.neubiorev.2021.12.029

Reynolds, C., Voress, J., & Pearson, N. (2005). *Test of Auditory Processing Skills-III* (3rd ed.). Proedinc Publishing.

Richard, G. J. (2017). *The source – Processing disorders* (2nd ed.). Pro-Ed.

Richard, G. J., & Hanner, M. A. (2005). *Language Processing Test, elementary.* Pro-Ed.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., Dahlström, Ö., Signoret, C., Stenfelt, S., & Pichora-Fuller, M. K. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7,* Article 31. https://doi.org/10.3389%2Ffnsys.2013.00031

Schow, R. L., Dillon, H., Hillam, J., Whitaker, M. M., & Seikel, J. A. (2021). Factor analysis on Multiple Auditory Processing Assessment–2 and Listening in Spatialized Noise–Sentences Test in children. *American Journal of Audiology, 30*(2), 433–442. https://doi.org/10.1044/2021_aja-20-00158

Sharma, M., Purdy, S. C., & Kelly, A. S. (2009). Comorbidity of auditory processing, language, and reading disorders. *Journal of Speech, Language, and Hearing Research, 52*(3), 706–722. https://doi.org/10.1044/1092-4388(2008/07-0226)

Smoski, W., Brunt, M. A., & Tannahill, J. C. (1998). *Children's Auditory Performance Scale (CHAPS).* Educational Audiology Association.

Steadman, M. A., & Sumner, C. J. (2018). Changes in neuronal representations of consonants in the ascending auditory system and their role in speech recognition. *Frontiers in NeuroScience, 12*, Article 671. https://doi.org/10.3389/fnins.2018.00671

Stephan (n.d.). *Restaurant ambiance* [Audio track]. Retrieved February 20, 2021, from https://soundbible.com/1664-Restaurant-Ambiance.html

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (4th ed.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231881.001.0001

Tomlin, D., Dillon, H., Sharma, M., & Rance, G. (2015). The impact of auditory processing and cognitive abilities in children. *Ear and Hearing, 36*(5), 527–542. https://doi.org/10.1097/aud.0000000000000172

Ueda, M., & Tanaka, T. (2020). *What is this sound in decibel? —We tried to determine students' comprehension of sound levels, online version* [Paper presentation]. 17th International Conference on Cognition and Exploratory Learning in Digital Age, Japan. https://eric.ed.gov/?id=ED626758

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (1999). *Comprehensive test of phonological processing: CTOPP.* Pro-Ed.

Warrier, C. M., Johnson, K. L., Hayes, E. A., Nicol, T., & Kraus, N. (2004). Learning impaired children exhibit timing deficits and training-related improvements in auditory cortical responses to speech in noise. *Experimental Brain Research, 157*, 431–441. https://doi.org/10.1007/s00221-004-1857-6

Webster, R. E. (2009). *Test of Information Processing Skills.* ATP Assessments.

Wilson, W. J., & Arnott, W. (2013). Using different criteria to diagnose (central) auditory processing disorder: How big a difference does it make? *Journal of Speech, Language, and Hearing Research, 56*(1), 63–70. https://doi.org/10.1044/1092-4388(2012/11-0352)

Yathiraj, A., & Maggu, A. R. (2013). Screening Test for Auditory Processing (STAP): A preliminary report. *Journal of the American Academy of Audiology, 24*(09), 867–878. https://doi.org/10.3766/jaaa.24.9.10

Yathiraj, A., & Vanaja, C. S. (2018). Criteria to classify children as having auditory processing disorders. *American Journal of Audiology, 27*(2), 173–183 https://doi.org/10.1044/2018_aja-17-0091

Zokoll, M. A., Fidan, D., Türkyılmaz, D., Hochmuth, S., Ergenç, İ., Sennaroğlu, G., & Kollmeier, B. (2015). Development and evaluation of the Turkish matrix sentence test. *International Journal of Audiology, 54*(sup2), 51–61. https://doi.org/10.3109/14992027.2015.1074735

Authors' Note

Correspondence concerning this article should be addressed to Hale Hancer, Department of Audiology, Faculty of Health Sciences, Baskent University, Baglica Campus, Ankara, 06810, TÜRKIYE. Email: **halehancer@gmail.com**

Acknowledgments

This research was supported by the Scientific and Technological Research Council of Turkey under the 1002 Quick Support Program (project number 121S261).

Disclosures

The authors declare no conflicts of interest, financial or otherwise. This manuscript was produced from a doctoral thesis titled "Development, Validity and Reliability of Auditory and Speech Performance Test for Children" (Ankara University, Institute of Health Sciences, 2023) by Hale Hancer.

Comparisons of Auditory Steady State and Auditory Brainstem Response Thresholds in Infants With Normal Hearing and Conductive Hearing Loss

Comparaison des seuils des réponses auditives à l'état stable et des potentiels évoqués auditifs du tronc cérébral chez les nourrissons ayant une audition normale et ceux ayant une perte auditive conductive

Hope Valeriote
Susan A. Small

KEYWORDS

AUDITORY BRAINSTEM RESPONSE (ABR)
AUDITORY STEADY-STATE RESPONSE (ASSR)
AIR CONDUCTION
BONE CONDUCTION
AIR-BONE GAP

Hope Valeriote[1,2,3] and Susan A. Small[3]

[1]Audiology Services, Glenrose Rehabilitation Hospital, Edmonton, AB, CANADA

[2]Early Hearing Detection and Intervention Program, Alberta Health Services, Edmonton, AB, CANADA

[3]School of Audiology and Speech Sciences, The University of British Columbia, Vancouver, BC, CANADA

Editor:
Victoria Duda

Abstract

This study investigated how well the air- and bone-conduction auditory steady-state response detected mild conductive hearing loss in young infants compared to the auditory brainstem response. Air-bone gap sizes were compared between infants with normal hearing and those with conductive loss using a two-group cross-sectional design. Twenty-three (500 Hz) and 22 (2000 Hz) infants (0–6 months of age) with normal hearing and 15 (500 Hz) infants with conductive loss were recruited from newborn hearing screening. Thresholds were obtained to frequency-specific air- and bone-conducted stimuli. There were no instances of conductive loss at 2000 Hz. Mean 500-Hz thresholds and air-bone gap sizes were compared. Sensitivity and specificity for identifying conductive loss were measured. Overall, mean bone-conduction thresholds were similar between groups, and mean 500-Hz air conduction thresholds were higher with larger air-bone gap size for infants with conductive loss. Sensitivity and specificity for identifying conductive loss were highest for air-conduction auditory brainstem response threshold measurement compared to screening and auditory steady-state response threshold measurements. Compared to the auditory brainstem response, the variability of auditory steady-state response thresholds and air-bone gap size was too great to reliably separate normal hearing from mild conductive loss. More research is needed using infants with varying degrees of hearing loss at multiple frequencies to fully assess the appropriateness of the auditory steady-state response as a clinical diagnostic tool for an infant population.

Abrégé

La présente étude a examiné dans quelle mesure les réponses auditives à l'état stable détectaient mieux les pertes auditives conductives légères chez les nourrissons que les potentiels évoqués auditifs du tronc cérébral, tant par conduction aérienne que par conduction osseuse. L'ampleur des écarts aériens-osseux entre les nourrissons ayant une audition normale et ceux ayant une perte auditive conductive a été comparée en utilisant un devis transversal à deux groupes. Vingt-trois (500 Hz) et vingt-deux (2000 Hz) nourrissons âgés de zéro à six mois ayant une audition normale et quinze (500 Hz) nourrissons ayant une perte auditive conductive ont été recrutés par l'intermédiaire d'un programme de dépistage de la surdité chez le nouveau-né. Les seuils de conduction osseuse et de conduction aérienne pour des stimuli de fréquences spécifiques ont été mesurés. Aucune perte auditive conductive n'a été observée à 2000 Hz. La moyenne des seuils à 500 Hz et l'ampleur des écarts aériens-osseux ont été comparés. La sensibilité et la spécificité des mesures recueillies pour la détection des pertes auditives conductives ont été calculées. De façon générale, la moyenne des seuils de conduction osseuse était similaire pour les deux groupes, tandis que la moyenne des seuils de conduction aérienne à 500 Hz était plus élevée et les écarts aériens-osseux étaient plus grands chez les nourrissons ayant une perte auditive conductive. La sensibilité et la spécificité pour la détection de la perte auditive conductive étaient plus élevées pour les potentiels évoqués auditifs du tronc cérébral par conduction aérienne lorsque comparées aux résultats du test de dépistage et aux seuils des réponses auditives à l'état stable. Lorsque comparée aux potentiels évoqués auditifs du tronc cérébral, la variabilité des seuils des réponses auditives à l'état stable et de l'ampleur des écarts aériens-osseux était trop grande pour distinguer fiablement une audition normale d'une perte auditive conductive légère. Davantage de recherches auprès de nourrissons ayant une perte auditive dont le niveau de sévérité varie et qui touche différentes fréquences sont nécessaires pour mieux évaluer la pertinence des réponses auditives de l'état stable à titre d'outil de diagnostic clinique chez les nourrissons.

Early hearing detection and intervention (EHDI) programs continue to be implemented internationally and are a major driving force for current research on infant hearing. Within this context, finding more efficient, less time-consuming methods of evaluating hearing in young infants has been of interest, with some looking for alternatives to the current diagnostic gold-standard method, the auditory brainstem response (ABR), to integrate in their programs. For decades, the ABR has been the proven and reliable method that major EHDI programs (Bagatto et al., 2020; Hatton et al., 2022; Joint Committee on Infant Hearing, 2019) rely on to identify hearing loss in the infant population. Like behavioural methods of auditory assessment, ABR thresholds can be obtained using both air- and bone-conducted stimuli (Stapells & Ruben, 1989; Yang et al., 1987, 1993). Bone-conduction (BC) ABR assessment of infant hearing is required to differentiate between conductive and sensorineural hearing losses when air-conduction (AC) thresholds are elevated (Hatton et al., 2012). When conductive hearing loss (CHL) is present, an air-bone gap (ABG) can be seen between elevated AC and normal BC thresholds.

Previous research to assess optimal testing conditions, including bone oscillator placement on the infant head and coupling method (i.e., handheld vs. band) and force, have informed clinical best practice in assessing BC thresholds in this young population who are unable to respond behaviourally (Small et al., 2007; Yang et al., 1991). The auditory steady-state response (ASSR) is another auditory evoked potential that is of interest to clinicians and researchers as an alternative to ABR because it also assesses hearing using both AC and BC stimuli, with the added benefit of testing multiple frequencies and ears simultaneously.The decision to use the ASSR as a method to assess infant hearing requires clear demonstration that the ASSR is comparable (or superior) to current gold standards in its ability to assess infant AC and BC frequency-specific hearing thresholds in an accurate and precise way.

An EHDI program requires high sensitivity to detect hearing loss in its target population and high specificity in differentiating elevated from normal hearing (NH) levels. Comparisons are needed between the ASSR and current gold-standard methods that take into account the participants' age (i.e., behavioural audiometry as the gold standard for infants > 6 months and ABR for infants ≤ 6 months; Gorga et al., 2006; Joint Committee on Infant Hearing, 2019; Widen et al., 2005), AC and BC modes of presentation, and hearing presentations from NH ability to all degrees of sensorineural, conductive, and mixed hearing loss. Maturational changes in BC ASSR responses and infant/adult differences in skull properties were discussed in Small and Stapells (2008a). Given these differences, it is important to understand how the BC ASSR behaves in infants with hearing loss. External ear canal changes and maturation for AC are understood and can be measured, but maturation effects for these infants are not as well documented for BC results. Understanding this is relevant to the use of ASSR/ABR in the clinical setting.

A review of the literature showed that many of these comparisons have been investigated and published over the last few decades, and comparisons between ASSR and behavioural thresholds to date have been encouraging. For example, several studies have shown that AC and BC ASSR and behavioural thresholds in infants with NH (Casey & Small, 2014; Luts et al., 2006) and infants with hearing loss (Aimoni et al., 2018) correlate highly. One study with a small number of participants also suggested the BC ASSR was able to identify normal cochlear sensitivity in young children with CHL (Nagashima et al., 2013).

It is known that there are differences in BC ASSR thresholds by age, likely due to skull maturation (discussed in detail in Casey & Small, 2014, and Small & Stapells, 2008a); however, the ASSR introduces other possible factors to consider, such as the use of high stimulus rates and multiple stimuli, that may have an effect on ASSR thresholds that does not apply to the ABR. It is especially important to establish a comprehensive body of literature for infants who are too young to respond behaviourally and who define the target population for EHDI programs.

Studies comparing AC ASSR thresholds to the tone-ABR (using varied stimulus parameters and test protocols) thus far have shown that AC ASSR thresholds (in dB HL) in infants with NH and with hearing loss are consistently poorer than AC ABR thresholds (in dB normalized hearing level [nHL]) but are highly correlated and accurate (Michel & Jørgensen, 2017; Rance & Rickards, 2002; Rance et al., 2006; Rodrigues et al., 2010; Van Maanen & Stapells, 2009, 2010). More recently, however, at least one other study using different collection protocols and stimulus parameters has suggested the reverse, with thresholds (in dB estimated hearing level [eHL]) for ABR being poorer than for ASSR (Sininger et al., 2018) This latter study, however, used larger correction factors for ASSR compared to ABR, which may explain that finding. More studies with infants with hearing loss are needed, but the existing AC ASSR data appear promising. Studies that compare frequency-specific BC toneburst-ABR thresholds with sinusoidal amplitude modulated (SAM) tone ASSR thresholds in young infants with NH and hearing loss are significantly lacking, thus this is the focus on the present study.

Small and Stapells (2008a) published BC ASSR normative data proposing a set of "normal" or minimum ASSR intensities, among others, for infants 0 to 11 months but did not compare frequency-specific BC ASSR to BC ABR. Swanepoel et al. (2008) did provide some BC ASSR data for infants with hearing loss but did not compare these results to toneburst-ABR results and tested a broad age range (0.25–11.5 years of age). To our knowledge, no data comparing BC ASSR to BC ABR thresholds in young infants with hearing loss exist in peer-reviewed publications.

For the present study, AC and BC ABR and ASSR thresholds in NH infants and infants with CHL confirmed by the gold-standard ABR thresholds were compared to investigate the following questions:

1. What are ABR and ASSR thresholds in young infants with NH and CHL?
2. How does the ABG compare between ABR and ASSR in young infants with NH and CHL?
3. What are appropriate minimum intensity cutoffs to differentiate NH from CHL using AC and BC ASSR in young infants?
4. Does the ASSR detect CHL as well as the ABR does in young infants?

Methods

Stimulus and recording setups for ABR followed those described in the British Columbia Early Hearing Program (BCEHP) ABR protocol (Hatton et al., 2022). Stimulus and recording setups for ASSR were similar to those described in Casey and Small (2014), with minor differences between the research and clinical versions of the MASTER software. Details of methodology for ABR and ASSR are provided below.

Participants

Infants were recruited through the newborn hearing screening program at the Royal University Hospital, Saskatoon. Participants were recruited if they failed newborn hearing screening or if they were unable to be screened at birth. Participation was entirely voluntary. Sixty-four infants between the ages of 0 and 6 months participated (NH mean age = 7.36 weeks, range 0.6–12.9 weeks; CHL mean age = 6.71 weeks, range 2.9–20.6 weeks); 61 from the well-baby nursery and 3 graduates from the neonatal intensive care unit, none of whom presented with congenital aural atresia or microtia. Fourteen infants were excluded because they did not sleep and did not complete any conditions of the testing session.

Each frequency (500 and 2000 Hz) was assessed individually and was categorized as a NH or CHL threshold based on the relationship between the AC and BC ABR results. In other words, if AC ABR was within normal limits at a specific frequency, that frequency's threshold was placed in the NH group. If AC ABR was elevated with normal BC ABR results at the same frequency, the frequency's threshold was placed in the CHL group. Normal versus elevated levels correspond to those specified in the BCEHP protocols. The BCEHP minimum stimulus intensities for 500 Hz are for AC, 35 dB nHL and for BC, 20 dB nHL; and for 2000 Hz are for AC, 30 dB nHL and for BC, 30 dB nHL (Hatton et al., 2022).

Results were included in the analysis whether partial or complete conditions were obtained. Infants who did not complete any portion of the protocol due to inability to sleep were excluded.

To verify the status of the middle ear and hearing at the time of testing, 1000 Hz tympanometry and transient-evoked otoacoustic emissions (TEOAEs) were performed using a Madsen AccuScreen and the OTOflex. The primary purpose of the screening measures was to corroborate the presence of middle ear pathology when participants were identified with CHL shown by abnormal tympanograms and absent TEOAEs and to determine the follow-up protocol per the Royal University Hospital guidelines. The cross-check principle has been used in pediatric audiology for decades. As Hall (2016) described, "no auditory test result should be accepted and used in the diagnosis of hearing loss until it is confirmed or crosschecked by one or more independent measures" (p. 59). TEOAE stimulus levels ranged from 70 to 84 dB SPL and used noise-weighted averaging. The response detection method involved the counting of significant signal peaks with self-calibration depending on ear canal volume. To pass the TEOAE test, a total of eight valid peaks in alternating directions (counted both above and below the median line) must be present.

Of the 50 participants who completed the testing, 31 did not pass tympanometry and 31 did not pass OAEs. A tympanogram was considered to be a "refer" if there was no identifiable peak or maximum admittance was less than or equal to 0.6 mmho compensated from the negative tail at −400 daPa, and thus, the tympanogram was considered flat (type B). Type B tympanograms (Jerger, 1970) show minimal or no mobility of the tympanic membrane supportive of otitis media with effusion and are considered an abnormal tympanometric pattern. TEOAEs were considered to be a "refer" if there was a response in fewer than three bands. Data collection took place in the context of a single audiology visit per participant, and the results of any medical and/or audiological follow-up is unknown.

Stimuli

AC stimuli were presented to participants using an ER-3A insert earphones in one ear (the same ear that was used to establish BC thresholds). BC stimuli were presented to participants using the B-71 bone oscillator placed on the mastoid, slightly posterior to the upper portion of the pinna for both ABR and ASSR testing. Small et al. (2007) showed no difference between lower and upper mastoid bone oscillator placement, so the upper portion was chosen to avoid interfering with the nearby mastoid electrode. This was coupled to the head with approximately 400 grams of force using the hand-held method (i.e., held by the first author). This coupling method was used as it was the least disruptive method to the infants' sleep and was found to have no significant differences to thresholds obtained by the elastic headband coupling method (Small et al., 2007). The examiner was trained to apply 400 grams of force (425 ± 25 g) by practicing BC application on a compressive spring scale, pressing down on the transducer with one or two fingers until the desired force was achieved with feedback. Once trials were completed with feedback, additional trials were completed without feedback, in a method similar to Small et al. (2007). Once it was determined that examiner was adequately trained, data collection began. Force was not verified on the infant head during testing. For ABR testing, the Intelligent Hearing Systems (IHS) SmartEP was used to generate and present stimuli. BCEHP-specified stimuli were used for both 500 and 2000 Hz (Hatton et al., 2022). These stimuli were exact-Blackman-windowed tones (five-cycle total duration, no plateau) and presented at a rate of 39.1/s to one ear (Hatton et al., 2019, 2022; Janssen et al., 2010). For ASSR testing, the two-channel Master II Clinical System was used to generate and present ASSR stimuli with carrier frequencies 500, 1000, 2000 and 4000 Hz. These stimuli were amplitude-modulated tone (AM2) at modulation frequencies 78, 85, 93 and 101 Hz for carrier frequencies 500, 1000, 2000 and 4000 Hz, respectively, and were presented simultaneously to one ear (monotic multiple [MM] ASSR).

Calibration

ABR Stimuli

AC stimuli were calibrated in dB nHL using ppeSPL with a Quest 177 sound level meter and G.R.A.S. DB 0138 2-CC coupler with 1-inch microphone. The acoustic calibrations for 0 dB nHL for AC using insert earphones, were 22 and 20 dB ppeSPL for 500 Hz and 2000 Hz, respectively. BC stimuli were calibrated using the B & K 4930 artificial mastoid, where the acoustic calibration for 0 dB nHL for BC using the B-71 bone oscillator were 67 and 49 dB re: 1μN ppe at 500 and 2000 Hz, respectively (see Stapells & Small, 2017; BCEHP, Hatton et al., 2022; or Ontario Infant Hearing Program, Bagatto et al., 2020 protocols for Canadian ppeRETSPLs and ppeRETFLs).

ASSR Stimuli

AC ASSR stimuli were calibrated in ppe SPL using the dB SPL RETSPLs per the American National Standards Institute (ANSI, 1996) using a Quest 177 sound level meter and G.R.A.S. DB 0138 2-CC coupler with 1-inch microphone. Each of the four frequencies were calibrated separately in dB HL and then combined. Calibrations for 0 dB HL for AC using insert earphones were 5.5 and 3 dB SPL for 500 Hz and 2000 Hz, respectively (ANSI, 1996). BC stimuli were similarly calibrated using (ANSI, 1996) RETFLs with the Quest 177 sound level meter and B & K Mastoid 4930 artificial mastoid. Calibrations for 0 dB HL were 58 and 31 dB re: 1μN for BC at 500 and 2000 Hz, respectively.

Recording

All participants were tested at the Royal University Hospital, Saskatoon in a double-walled sound-attenuating booth. All recordings were obtained using the Intelligent Hearing Systems SmartEP ABR system and the Master II Natus/Biologic clinical ASSR System. Four disposable electrodes were placed on the infant's scalp using the typical electrode montage for infant ABR testing: one (non-inverting) electrode on the vertex, an (inverting) electrode on each mastoid and the common electrode off-center on the forehead. Impedance for each electrode was less than 3 kOhms.

For ABR testing, standard BCEHP parameters were used at the time of data collection. Gain was set to 100,000 and band-pass filtering from 30 to 1500 Hz with an artifact rejection of plus or minus 25 μV. One channel was recorded for AC ABR conditions and two channels were recorded (i.e., the ipsilateral and contralateral montages) for BC ABR conditions. A minimum of two replications of 2000 trials each was obtained at threshold levels and one step (of 10 dB) below threshold. The presence and/or absence of a response were determined visually and by objective measures (signal to noise ratio and residual noise) and was interpreted by the first author. A response being "present" was determined by a visually identifiable wave V in the averaged waveform in the ipsilateral channel recording for AC ABR and by an ipsilaterally dominant wave V response for BC ABR when ipsilateral and contralateral recordings were compared (Hatton et al., 2022).

In accordance with BCEHP guidelines, "no response" was determined only when no visually identifiable wave V was present and SmartEP residual noise was less than or equal to 0.08 μV (Hatton et al., 2022). For ABR measures, where a response was identified, RN and SNR measures were used to support visual interpretation where possible, but ultimately visual identification of a response was considered sufficient to determine "response present." "No response" judgements were made on the basis of both visual interpretation but also with IHS-SmartEP SNR values less than 1 and RN less than or equal to 0.08μV. The participants were classified as having CHL at a frequency by demonstrating elevated AC ABR results with normal BC ABR results (based on BCEHP levels) with abnormal tympanometry findings.

In the classification of participants, tympanometry and OAE screening was used only as a cross check to confirm CHL where identified at a specific frequency to provide additional evidence in the identification of CHL. The NH group did not have a specific criterion for tympanometry or OAE screening result and that categorization was made on the basis of present ABR to AC and BC stimuli at minimum normal levels at that frequency (Hatton et al., 2022). To ensure the validity in this method of categorization, an independent samples *t* test was performed comparing AC ABR and ASSR NH thresholds in infants with normal OAE screening results and abnormal OAE screening results for 500 Hz and 2000 Hz separately and was not significant. This supports the validity of this method of categorization. Where the typical clinical protocol used in BCEHP does not include testing down to threshold if a present response has been established at the minimum stimulus intensity, testing down to threshold for all measures completed (AC and BC ABR and ASSR) did take place in this study.

For ASSR testing, two channels were recorded but only the ipsilateral channel (i.e., vertex-ipsilateral mastoid) was examined when determining response presence or absence and was the only channel analyzed in this study. For ASSR measures, no visual identification was required, as only SNR (*p* value) and residual noise were used to make response/no-response determinations. Masking was not used as the interaural attenuation for infants reported by Small and Stapells (2008b) is at least 10 to 30 dB. The EEG was filtered using a 30 to 150 Hz filter and amplified 10,000 times with artifact rejection set to plus or minus125 μV. The analog-to-digital conversion rate was 1200 Hz. Each sweep consisted of 16 epochs of 1024 data points and took 13.11 seconds of recording time. The ASSRs were averaged in the time domain and analyzed online in the frequency domain using a fast Fourier transform with a resolution of 0.08 Hz over a range of 0 to 625 Hz. Amplitudes were measured baseline-to-peak and expressed in nV. Recording continued until there was a response present with a minimum of 10 sweeps, or the residual noise levels were at least less than 15 nV and there was a minimum of 10 sweeps completed; whichever came first. An *F* ratio was calculated by the MASTER II system and a response was considered present if a significant response value ($p < .05$), was obtained from the F ratio compared to critical values for $F(2, 240)$ for at least three consecutive sweeps. The F ratio estimated the probability that the amplitude of the ASSR at the modulation frequency was significantly different from the average amplitude of the noise at adjacent frequencies. This was calculated within 120 bins, or plus or minus 60 bins from the modulation frequency (John & Picton, 2000). A response was considered absent if no significant response value was obtained ($p > .05$) and the noise value was appropriately low (< 15 nV).

Procedure

One session lasting between one and three hours took place for each subject. Before testing began, caregivers consented to participation in the study and were provided a small honorarium. All infants completed a hearing screening (TEOAEs), ABR and ASSR testing in the recording session. One ear was chosen to be tested using electrophysiologic methods. The selection of test ear was made based on the outcome of the hearing screening. If only one ear failed, that ear was tested using ABR and ASSR; if both ears failed, or both ears passed, the ear was chosen based on the most comfortable position for the infant and the caregiver. Testing was completed with the examiner inside the booth, next to the infant and caregiver. The examiner held the oscillator and continually monitored the placement of the earphone in the infant's ear.

Hearing screening using TEOAEs and 1000 Hz tympanometry was conducted and electrodes were applied while the infant was awake. The infant was given the opportunity to fall asleep before ABR and ASSR testing began and remained asleep during these tests in the caregiver's arms during ABR and ASSR testing. If the infant woke during the session, an opportunity for them to fall back of sleep was given before testing continued. Electrophysiological testing always began with ABR in order to provide parents with information from a gold-standard test before proceeding with ASSR. AC ABR was followed by BC ABR.

ABR testing began at BCEHP minimum stimulus intensities that correspond to the upper limit of NH. For AC ABR, testing began at 35 dB nHL and 30 dB nHL for 500 Hz

and 2000 Hz, respectively. BC ABR testing began at 20 dB nHL and 30 dB nHL for 500 and 2000 Hz, respectively. No masking was used given the age group of the participants and large interaural attenuation. A 10-dB bracketing method was used. Threshold levels were defined as the lowest level at which a response is present with an absent response 10 dB below. The lowest level tested was 0 dB nHL for 2000 Hz AC and BC and 500 Hz BC, and 5 dB nHL for 500 Hz AC due to starting levels and 10-dB step sizes.

ASSR testing began at 30 dB HL, corresponding to the highest "minimum level" obtained by Casey and Small (2014). Similar to the ABR testing procedure, a 10-dB bracketing method was used, and threshold was defined as the lowest level at which a response is present with an absent response 10 dB below. The lowest level tested was 0 dB HL for ASSR. As the study was more interested in BC comparisons, BC ASSR was prioritized over AC and was completed first. Thresholds were found using a 10-dB bracketing procedure. If a response was present, intensity was decreased by 10 dB. If no response was present, intensity was increased by 20 dB. Testing continued down to threshold.

In the sections to follow, only 500-Hz data for NH and CHL groups are discussed. Although 2000-Hz data were collected, as there were no instances of CHL observed at 2000 Hz, these data are only briefly addressed in the remainder of this article.

Data Analyses

AC and BC ABR and ASSR Thresholds

Measures of central tendency and dispersion are provided for each frequency condition (500 & 2000 Hz) by group (NH, CHL). Independent samples *t* tests were performed to compare 500 Hz AC and BC ABR and ASSR thresholds between NH and CHL groups.

ABGs

ABG between ABR and ASSR in infants with NH and CHL were compared. ABGs were calculated for ABR by subtracting the BC (nHL) threshold from AC (nHL) threshold for each subject. ASSR ABGs were calculated by subtracting the BC (dB HL) threshold from the AC (dB HL) threshold. For 500 Hz, an independent samples *t* test was then conducted to determine if means between groups were significantly different. Differences in thresholds were considered significant at the $p < .05$ level.

Minimum Normal Intensities

Individual and mean AC and BC ABR and ASSR thresholds were determined for NH (500 and 2000 Hz) and CHL (500 Hz) groups. "Minimum normal intensities" represent the minimum test intensity a response would need to be present in a clinical setting to confirm NH. Minimum intensities were determined by calculating the cumulative percent of responses present at each stimulus level for each testing method and mode of presentation. The intensity at which greater than 90% of NH infants had a response was considered the "minimum intensity" (e.g., Small & Stapells, 2008a; Van Maanen & Stapells, 2009). Ninety percent was chosen to represent an intensity that separates "normal" from "elevated" well according to the gold-standard threshold measure, ABR. The intensities for BC ABR have been assessed by Hatton et al. (2012).

Sensitivity and Specificity for CHL Detection With ASSR

Sensitivity and specificity using ASSR was measured using the ABR as the gold standard to determine whether the ASSR detects CHL as well as the ABR. Sensitivity and specificity of the OAE/tympanometry screening, AC ASSR threshold and ABG and ABR ABG were also measured. These were calculated as shown in **Table 1**. The 500 Hz ABR and ASSR thresholds were averaged separately for AC and BC, and 500 Hz thresholds were compared between normal and CHL groups. Analyses were performed using an independent samples *t* tests. Differences in thresholds were considered significant at the $p < .05$ level.

Results

Mean thresholds (as well as SD and 90% levels) for both 500 and 2000 Hz ABR/ASSR AC and BC thresholds can be found in **Table 2**. For infants with confirmed CHL, mean AC ABR thresholds increased compared to infants with NH. Mean AC ABR and ASSR thresholds at 500 Hz were larger for infants with CHL. Standard deviations were greater for ASSR thresholds compared to ABR thresholds for both AC and BC. No infants demonstrated CHL at 2000 Hz (as defined as an elevated 2000 Hz AC threshold in the presence of a 2000 Hz BC threshold within normal limits) and for this reason are not discussed in sections to follow.

500 Hz ABR and ASSR Thresholds for CHL and NH Groups

Independent samples *t* tests showed AC ABR thresholds for the CHL group were higher than thresholds for the NH group[$t(36) = -10.95$, $p < .001$]. BC thresholds did not significantly differ across groups [$t(37) = -0.67$, $p = .51$]. AC ASSR thresholds for the CHL group were also higher than thresholds for the NH group [$t(34) = -2.10$, $p = .043$]. BC ASSR thresholds did not significantly differ across groups [$t(36) = 0.56$, $p = .579$]. Levene's Test for equality of variances was not significant, so equal variances were assumed. AC ABR and ASSR and BC ASSR

Table 1

Sensitivity and Specificity Calculation Method

Condition	CHL on ABR	NH on ABR
CHL on ASSR	A = True positive	B = False positive
NH on ASSR	C = False negative	D = True negative

Note. Sensitivity = A/(A+C) x 100, specificity = D/(D+B) x 100. ASSR = auditory steady-state response; ABR = auditory brainstem response; NH = normal hearing; CHL = conductive hearing loss.

Table 2

AC and BC 2000 and 500 Hz ASSR and ABR Mean Thresholds for NH and CHL Groups

Measure	NH group				CHL group	
	2000 Hz		500 Hz		500 Hz	
	ASSR (dB HL)	ABR (dB nHL)	ASSR (dB HL)	ABR (dB nHL)	ASSR (dB HL)	ABR (dB nHL)
AC						
M	20.47	17.72	29.52	25.43	36.67	48.33
SD	12.03	9.22	9.20	7.05	11.12	4.88
n	21	22	21	23	15	15
90% level	40	30	40	35		
BC						
M	21.00	15.00	17.39	10.41	15.33	12.00
SD	13.96	10.12	9.63	7.50	12.63	7.93
n	21	22	23	24	15	15
90% level	40	30	30	20	40	20

Note. AC = air-conduction; BC = bone-conduction; ASSR = auditory steady-state response; ABR = auditory brainstem response; NH = normal hearing; CHL = conductive hearing loss; 90% level = lowest level (in dB HL or nHL) at which at least 90% of group showed response present.

and ABR thresholds were not correlated [AC slope = 0.17, intercept = 26.67, R^2 = 0.04; BC slope = 0.15, intercept 15.35, R^2 = 0.01].

ABG for CHL Versus NH Groups

The ABGs at 500 Hz for the groups with NH and CHL are shown in **Table 3**, **Figure 1** (ABR), and **Figure 2** (ASSR). On average, the mean 500-Hz ABGs for the two groups were as follows: (a) ABR NH, 15 dB, (b) ASSR NH, 12 dB (c) ABR CHL, 36 dB, and (d) ASSR CHL, 21 dB. Two outliers were present in the ABR ABG data set for the CHL group. The independent samples bootstrapped *t* test comparing mean ABR ABG for the NH and CHL groups showed significantly larger ABGs for the CHL group than in the NH group [$t(35) = -7.74$, $p < .001$]. Outliers were defined as a data point greater than the upper quartile plus 1.5 times the interquartile range. A bootstrapped *t* test was used to account for the outliers. Similarly, for the ASSR, the independent samples *t* test showed ABGs were also significantly larger in CHL group [$t(33) = -2.30$, $p = .028$]. Levene's Test for equality of variances was not significant, so equal variances were assumed. The majority of NH participants had 500-Hz ABR ABGs 25 dB or smaller (21 of 22) and CHL subjects of 35 dB or larger (13 of 15). As shown in **Figures 1** and **2**, compared to ABR, there was more overlap between ASSR NH and CHL groups where the majority of NH participants had ASSR

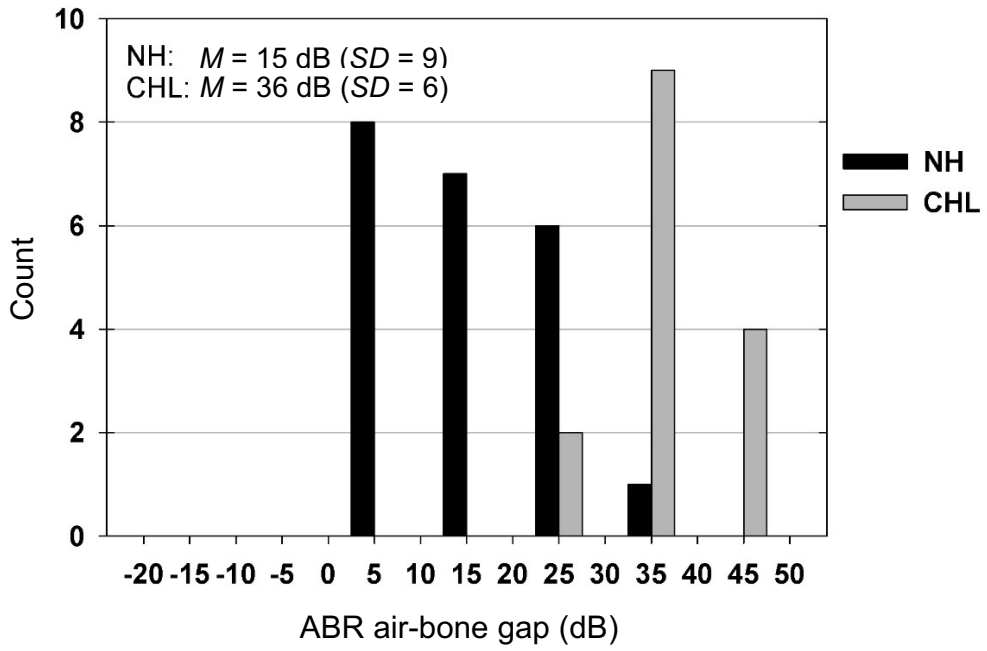
Table 3

2000 and 500 Hz ASSR and ABR Mean ABG Size for NH and CHL Groups

Measure	NH group				CHL group	
	2000 Hz		500 Hz		500 Hz	
	ASSR ABG (dB)	ABR ABG (dB)	ASSR ABG (dB)	ABR ABG (dB)	ASSR ABG (dB)	ABR ABG (dB)
M	−3.33	2.73	12.00	15.00	21.33	36.33
SD	15.92	12.41	10.56	9.26	13.56	6.40
n	21	22	20	22	15	15

Note. ASSR = auditory steady-state response; ABR = auditory brainstem response: ABG = air-bone gap; NH = normal hearing; CHL = conductive hearing loss.

Figure 1



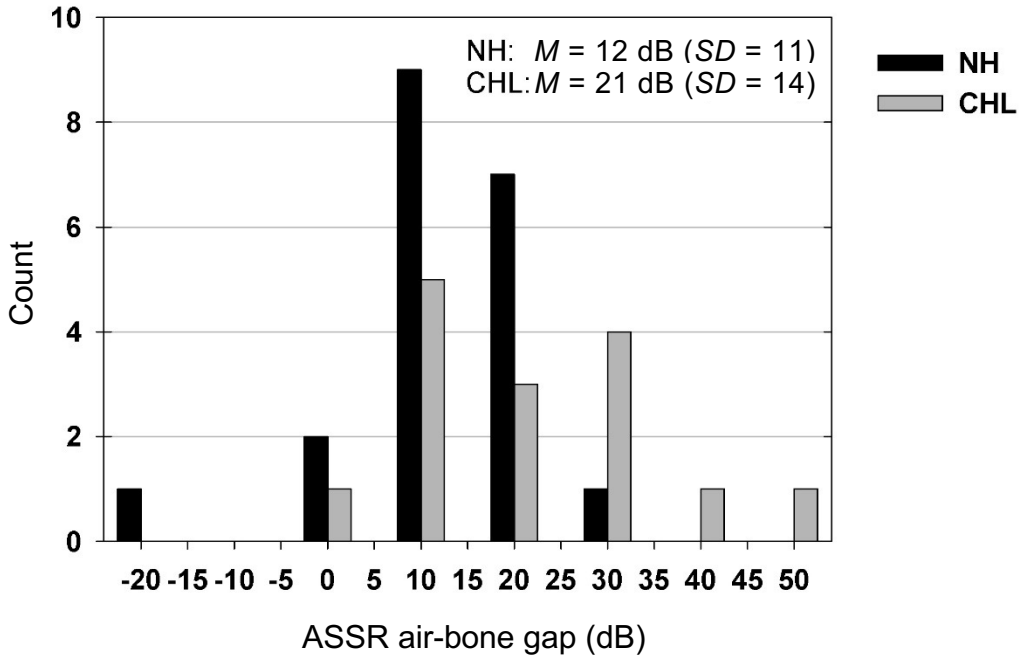
Auditory brainstem response (ABR) air-bone gaps in normal hearing (NH) and conductive hearing loss (CHL) groups

ABGs of 20 dB or smaller and CHL of 10 dB or larger. ABR and ASSR ABG size, however, were poorly correlated (slope = 0.24, intercept = 10.02, R^2 = 0.07). For 2000 Hz, the mean ABG for the NH group was approximately −3 and 3 dB for ASSR and ABR, respectively; the majority of NH infants had 2000-Hz ABR ABGs 10 dB or smaller (20 of 22), and 10 dB or smaller for ASSR (19 of 21).

Minimum "Normal" Intensities

The minimum intensity cutoffs for ABR and ASSR were determined for AC and BC separately by calculating the cumulative percent of responses present at each stimulus intensity for each testing method and mode of presentation. The intensity at which greater than 90 percent of NH infants had a response was considered the "minimum intensity."

Figure 2



Auditory steady-state response (ASSR) air-bone gaps in normal hearing (NH) and conductive hearing loss (CHL) groups

For ABR, the minimum intensity levels were 30 dB nHL for 2000 Hz AC and BC, and 35 and 20 dB nHL for 500 Hz AC and BC, respectively (see **Table 2**). For ASSR, the minimum intensity levels were 40 dB HL for 2000 Hz AC and BC, and 40 and 30 dB HL for 500 Hz AC and BC, respectively. Minimum ASSR intensities (in dB HL) were found to be higher for AC compared to BC and higher for ASSR compared to the ABR minimum intensities (in dB nHL) used by the BCEHP.

Sensitivity and Specificity for CHL Detection

Table 4 shows a comparison of sensitivity and specificity for each test completed using different cutoff criteria. Sensitivity and specificity were calculated, with reference to NH or CHL diagnosis using the gold standard ABR, for Tympanometry/OAE screening, ASSR and ABR ABG size, and AC ASSR thresholds. For such a test to be used in the context of an EHDI program, there is very low tolerance for missing cases of hearing loss, so sensitivity must be high. For screening measures, sensitivity was high for all screening tests, while specificity was poor. For ABG, the sensitivity worsens and specificity improves as the criteria to define the minimum ABG for CHL increases. Using AC ASSR thresholds as the criterion for CHL identification, as threshold increases, sensitivity worsens while specificity improves.

Discussion

ABR Versus ASSR Threshold Differences

ABR and ASSR thresholds differed when measured in the same infant and were found to be poorly correlated in the present study (AC r = .04; BC r =.01). Given the narrow range of thresholds obtained, this is not surprising. Previous studies have shown strong correlations between ABR and ASSR thresholds, and they are thought to measure responses from approximately the same part of the auditory system (Sininger et al., 2018; Van Maanen & Stapells, 2010). For these reasons, it was expected that the ABR and ASSR thresholds would be similar. Threshold differences in the present study may be in part due to the 0 dB HL/nHL values used for ABR and ASSR, starting intensities being offset by 5 dB (i.e., ABR at 35 dB nHL and ASSR 30 dB HL), a 10-dB step size, and the rather modest AC threshold shifts due to mild CHL. If smaller step sizes, similar starting levels, and populations with greater degrees of CHL were used, this difference may have followed the trend of Rance et al. (2006) where the difference between thresholds became minimal when converted to like units. Other contributors to the ASSR-ABR differences are potentially related to an insufficiently large sample size and a difference in stimuli.

Table 4

Sensitivity and Specificity of CHL Detection Compared to Gold Standard ABR by Test Type

Category	Test	Sensitivity	Specificity
Screening	Tymp + OAE	100 [a]	44
	OAE only	100 [a]	58
	Tymp only	**100** [a]	**60**
ASSR ABG size (dB)	≥ **10**	**93**	**15**
	≥ 20	60	60
	≥ 30	40	95
	≥ 40	13	100
ABR ABG size (dB)	≥ 5	100	0
	≥ 15	100	36
	≥ **25**	**100**	**68**
	≥ 35	87	95
	≥ 45	26	100
ASSR thresholds (dB HL)	**AC ≥ 20**	**100**	**5**
	AC ≥ 30	87	33
	AC ≥ 40	47	67
	AC ≥ 50	33	100

Note. CHL = conductive hearing loss; ABR = auditory brainstem response; ASSR = auditory steady state response; Tymp = tympanometry; OAE = otoacoustic emissions; ABG = air-bone gap; AC = air conduction. The subtype for each test category that had the highest specificity and a sensitivity exceeding 90% is bolded.
[a] Sensitivity of 100% due to classification criteria for CHL that required "refer" results on both tympanometry and OAE screening.

ABGs

To our knowledge, this is the first study to compare AC and BC ASSR thresholds within participants in infants with ABR-confirmed NH and CHL. ABGs for both ABR and ASSR CHL groups were larger than their NH counterparts, as expected. We had anticipated that threshold differences would be observed between ASSR and ABR thresholds because dB nHL threshold values were used for ABR (i.e., thresholds that included a consideration of temporal integration issues) whereas the dB HL values for SAM tones were based on dB HL values for long-duration tones as specified in ANSI S3.6 (ANSI, 1996). However, we also anticipated the differences in threshold for ASSR and ABR to be similar for AC and BC stimuli, thus no impact to estimated ABGs was expected (i.e., whatever differences may be present between ABR and ASSR would affect AC and BC thresholds similarly and thus the ABG would not be greatly affected across these measures). ABR and ASSR ABGs for the NH group did not differ significantly, however CHL ABR and ASSR ABGs did differ significantly.

We also expected that the ABGs for ABR and ASSR would be larger in infants with CHL than with NH and this was confirmed. In clinical practice when using behavioural methods of assessment, clinicians operate under the assumption that individuals with NH and sensorineural hearing loss do not exhibit clinically significant ABGs; while those with CHL are expected to show an ABG. In adult audiometry, with test-retest reliability in behavioural audiometry of plus or minus 5 dB, an ABG greater than or equal to 15 dB is often considered clinically significant, and this tends to be extrapolated in clinical practice to define clinically significant ABG sizes in pediatric assessments.

In the clinical setting, it is challenging to assess the magnitude of the ABR and ASSR ABGs for several reasons. First, most ABR/ASSR clinical protocols do not encourage testing down to true threshold (at least, not in Canada), but rather recommend the use of "minimum normal intensities" where if a response is present, it is considered within normal limits and testing at lower presentation intensities is not required. The goal of most Canadian EHDI programs is to detect permanent congenital hearing losses greater than or equal to 30 dB HL and this method of assessment accomplishes this goal in a time-effective manner.

Second, ABGs for clinical use (in many provinces in Canada) are calculated on the eHL values for ABR after nHL-to-eHL correction factors have been applied. The purpose of these frequency- and mode-specific correction factors is to more closely estimate pure-tone behavioural (dB HL) thresholds used for diagnostic and hearing aid fitting purposes (see the most current BCEHP clinical protocol for up-to-date correction factors, corresponding stimulus parameters and recording techniques; Hatton et al., 2022). When AC and BC eHL correction factors are applied, they come with their own estimation errors (as much as 10–20 dB of error in either direction), and these errors are additive when calculating the ABG. These estimation errors in combination with using 10-dB step sizes and/or not testing down to a true threshold make ABG estimations in clinical practice challenging. The ABR or ASSR ABG can be substantially under- or over-estimated, and therefore are more appropriately used in a descriptive way to comment on the size of a conductive component rather than a singular diagnostic criterion. It is important that clinical protocols recognize these limitations. Clinicians need to keep in mind that unique correction factors are applied to each individual frequency and differ for AC and BC. When investigating the diagnostic power of ABG size, this can apply only to a specific frequency and cannot not be generalized beyond the frequency that is being investigated. Carefully determined correction factors are necessary and will affect any measure of the ABG. More data are needed for each frequency with different degrees of CHL. This study continues to support the value of the ABG as a descriptive tool to accompany frequency-specific ABR thresholds and tympanometry measures in differentiating NH from CHL.

NH Thresholds and Minimum Intensities

As shown in **Tables 5** and **6**, across several studies, normal ASSR intensity levels (i.e., ASSR threshold level upper limits for those infants with no hearing loss) are approximately 50 and 40 dB HL for 500 and 2000 Hz AC, and approximately 20 to 30 and 40 dB HL for BC stimuli. The present study showed minimum intensities of 40 dB HL for both 500 and 2000 Hz AC and 30 and 40 dB HL for 500 and 2000 Hz BC for ASSR. The AC values differ by not more than 10 dB from the average of the other studies using different stimuli (note variations across studies in the modulation function, or use of one versus multiple simultaneous ASSR stimuli). These differences may be attributed to differences in sample size, stimuli, whether presentation was multiple or single, the age range tested, stopping criteria, EEG noise, recording system (and the system's detection algorithm). The BC minimum intensity at 500 Hz proposed in Small and Stapells (2008b) of 30 dB HL is consistent with what was found in the present study. The minimum intensity from this study was slightly higher than 20 dB HL reported by Casey and Small (2014), however, was likely somewhat overestimated (~5 dB) due

Table 5

A Summary of 500 & 2000 Hz AC ASSR Normal Levels in Young Children in the Literature

Study	Modulation	Multiple (M) / single (S)	Age	Norm max (dB HL)	
				500 Hz	2000 Hz
Lins et al. (1996)[a]	AM	M	1–10 months	48	38
Cone-Wesson et al. (2002)	AM	S	< 4 months	> 71	50
John et al. (2004)	MM, AM, AM^2	M	3–15 weeks	> 46	> 50
Rance et al. (2005)	MM	S	1–3 months	52	40
Swanepoel & Steyn (2005)	MM	M	3–8 weeks	50	> 50
Luts et al. (2006)[a]	MM	M	< 3 months	> 44	42
Rance et al. (2006)	MM	S	6 weeks	50	
Van Maanen & Stapells (2009)	AM (cos^3 sinusoids[b])	M	< 6 months	49	36
Casey & Small (2014)	AM^2	M	6.5–19.0 months	30	
Rodrigues & Lewis (2014)[c]	NB chirps	M	2 days	59	31
Present study	AM^2	M	0–6 months	40	40

Note. AC = air-conduction; ASSR = auditory-steady state response; AM = amplitude modulation; MM = mixed modulation; NB = narrowband; multiple and single refer to the amount of simultaneous ASSR stimuli.
[a] Thresholds were converted from dB SPL to dB HL using ANSI-1996 adjustment values.
[b] Cosine3-windowed sinusoids are nearly equivalent to AM^2.
[c] Levels provided are in dB HL, converted levels from NB chirp (nHL) using Haughton 2-cc coupler conversions.

Table 6

A Summary of 500 & 2000 Hz BC ASSR Normal Levels in Young Children in the Literature

Study	Modulation	Age (months)	Norm max (dB HL)	
			500 Hz	2000 Hz
Small & Stapells (2008)	MM	0–11 12–24	30 40	40 40
Small & Stapells (2008)	MM	2–11	30	30
Casey & Small (2014)	AM[2]	6.5–19.0	30	40
Present study	AM[2]	0–6	30	40

Note. BC = bone conduction; ASSR = auditory steady-state response; MM = mixed modulation; AM = amplitude modulation. All studies in this table used multiple simultaneous ASSR stimuli.

to the distribution of the threshold data and the step size used. The present study and that by Casey and Small (2014) are the only studies to date that used AM[2] stimuli for BC ASSR, and both studies showed better than the average of thresholds included in the table. However, ASSR amplitudes are in keeping with these studies and a study using AM/FM stimuli (Small & Stapells, 2008a). The mechanisms underlying this difference in thresholds remains an open question. In addition, maturation of the BC ASSR response makes "cutoff" points less clear for categorizing hearing loss and the use of an age range of 0 to 6 months may be too large. Minimum intensities for ABR were in keeping with those suggested by BCEHP. Importantly, they were not found to be lower. As mentioned earlier, *t* tests comparing NH participants' AC and BC ABR or ASSR thresholds at 500 and 2000 Hz between those with normal and those with abnormal OAE screening results showed no significant difference; thus, the NH group includes all NH thresholds, regardless of screening result. Minimum normal intensities for ABR also did not differ between those in the NH group with a normal versus abnormal OAE screen; both subgroups had minimum normal intensities of 30 for 2000 Hz AC and BC and 35 and 20 for 500 Hz AC and BC, respectively.

Sensitivity and Specificity for CHL Detection

As discussed earlier, 500-Hz AC ASSR thresholds are, on average, higher than BC ASSR thresholds in the CHL group, with a significantly larger ABG compared to the NH group (see **Tables 2** and **3**). In the case of mild CHL, however, the ABR does seem better at identifying modest ABGs in infants than the ASSR, using the parameters incorporated in this study. For this reason, clinicians should exercise caution when considering using ASSR AC/BC thresholds to identify mild CHL with small ABGs. Cutoff thresholds where the ASSR ABG demonstrated high sensitivity lacked high specificity (see **Table 4**). The sample for the current study did not include infants with sensorineural hearing loss, and therefore it was not possible to determine the number of false negatives for BC ASSR. There was significant overlap in the NH and CHL AC ASSR threshold distributions (hence the imperfect sensitivity and specificity). We hypothesize that this overlap may be a reflection of the sampling between groups. Overall, the CHL group only demonstrated a very mild degree of hearing loss that was isolated to 500 Hz (and perhaps in some cases may have been resolving). The NH sample recruited was primarily at-risk infants who failed or missed their initial hearing screening, and thus some subclinical degree of middle ear dysfunction may have been present. Perhaps if the CHL group demonstrated elevated AC thresholds across frequencies or the degree of loss was greater, and the NH group had instead been recruited from low-risk infants with confirmed normal middle-ear function, the overlap would have been minimized. Nevertheless, the present study operated under conditions that are typical in the clinical setting where the separation between these groups may be less than ideal.

Clinical Implications

Previous studies that compared ASSR thresholds to gold-standard methods of infant hearing assessment reported strong correlations between the methods and reasonable accuracy in estimating hearing thresholds. Some studies have suggested that the ASSR can accurately separate NH infants from those with hearing loss. Differing methodology in this body of research continues to be problematic. This study is one of only a small number of studies comparing the ASSR to tone-ABR in infants with NH and with hearing loss, and within these, methodologies and stimulus choices differ. More work still needs to be done to determine the best methodology and stimulus parameters for ASSR testing in infants. This

study added to the body of research by providing data using AC and BC ASSR to AM² tones at 500 Hz in infants with NH or CHL, but it would be beneficial to provide more data with different degrees of hearing loss and hearing loss at different frequencies. It is our opinion that the ASSR requires more research to better understand optimal test parameters for use as a screening or diagnostic measure for EHDI programs before being clinically implemented. When compared to the ABR and screening measures, ASSR thresholds underperform in their ability to detect mild low frequency CHL. The ABR is widely used, with well-studied diagnostic criteria and protocols and it is known that the ABR is accurate in differentiating NH sensitivity from a variety of degrees of CHL, mixed and sensorineural hearing loss in infants. At this time, the ABR continues to be the gold standard and is the diagnostic test method that clinicians should continue to use in their EHDI Programs.

Limitations and Future Directions

The study was conducted in the context of a clinical audiology department in Canada, where it is routine to have one individual making response judgements. This context was perhaps a limitation of the study. It should be noted that the same individual made threshold estimations for AC and BC ASSRs and ABRs, and hence any subjective bias across response measures should have been similar for all measures, and hence it is unlikely the choice of a single expert judge of response presence is a confounding factor in this study.

The moderate number of infants included in this study demonstrated a mild degree of CHL that was only observed at 500 Hz. The conclusions regarding the efficacy of the ASSR in detecting CHL could differ in a sample with more significant degrees of CHL, hearing loss that extends beyond 500 Hz, and in cases of more extensive middle ear dysfunction (e.g., congenital aural atresia, congenital fixation of the ossicular chain, acute otitis media). The results of the present study suggesting that the ASSR may not be as good an indicator of ABG and CHL (or middle ear abnormality) as the ABR may be limited to ASSR protocols using multitone AM² stimuli, and only for infants with what appears to be a rather mild conductive loss. Further studies with more participants and degrees of CHL may help to provide a larger picture of the efficacy of the ASSR in detection CHL in infants. Of note, acoustic reflexes were not measured in the initial assessment of infants in either NH or CHL groups. In future studies with a wider range of hearing loss degrees, the addition of acoustic reflexes may be beneficial to include to more completely assess middle ear status. Acoustic reflexes can be problematic, however, in that they may wake the infant and thus are a lower priority measure in most Canadian ABR protocols.

MM ASSR stimuli (500, 1000, 2000, 4000 Hz) were presented; however, due to time constraints only 2000 and 500 Hz were able to be measured down to true threshold and compared with ABR thresholds in this study. The assumption is that if this were to be used clinically as the only electrophysiologic measure of hearing, the clinician would ideally assess threshold to all four stimuli to provide a more complete threshold assessment. In previous studies, it has been demonstrated in human subjects that the use of multiple SAM stimuli is more efficient than a single SAM stimulus (Hatton & Stapells, 2011) and the present study aimed to investigate a possible alternative to the ABR using a technique that more time efficient than the single-frequency technique that is used in diagnostic ABR assessments. In this particular instance, focusing on two frequencies of interest (500 Hz, 2000 Hz) stimuli, while presenting four AM² stimuli may have added some noise to the recordings, and perhaps modestly elevated ASSR thresholds, but the authors are not aware of any published evidence that MM results in higher thresholds than MS in human subjects. It should be noted that four SAM² stimuli were used for both AC and BC ASSR stimuli, and hence would not have expected this to substantially influence the magnitude of the ABG. Future studies aimed to provide clinical evidence of threshold changes resulting from the use of multiple ASSR stimuli would directly address this possibility.

This study aimed to answer a research question while providing some clinically relevant information. As such the test protocol was inefficient and if adopted in a routine clinical setting, would almost certainly have resulted in limited or inadequate information being acquired before the infant wakes. The key procedural elements that would need to be changed in the clinical setting include the test strategy of testing down to true threshold for all infants (clearly unnecessary once it is known that their hearing is normal), as well as the use of both the ABR and ASSR as electrophysiological techniques to assess hearing in the same test session (not ever required clinically). The BCEHP protocol for clinical use recommends testing down to a minimum intensity rather than testing down to true threshold in infants with NH (Hatton et al., 2022). In addition, BCEHP does not include the measurement of ASSR thresholds in their clinical protocol; ASSR measures were made solely for the purposes of this study. Using only one electrophysiologic measure of hearing thresholds would significantly reduce test time and would allow for frequency-specific assessment of both ears in a time period that is more reasonable in a clinical setting.

An important additional procedural consideration is that ABR testing was the priority for the clinical portion of

the assessment, always taking place first. This may have had an impact on the noise of the later-recorded ASSR recordings. It is also worthwhile to explore test time and protocol efficiency of ASSR where there is a consideration for implementation in a clinical setting that is unique to the stimulus and recording parameters and protocols intended to be used (e.g., Cebulla & Stürzebecher, 2015; Sininger et al., 2018, 2020). The difference in starting levels (5-dB offset) between ABR and ASSR may have influenced the differences in ABG size between the two methods. Once correction factors for ASSR are confidently established, a different starting level for ASSR may be more appropriate.

Finally, due to the properties of the underdeveloped skull, the threshold for BC stimuli at 500 Hz is lower (better) in infants than in adults (e.g., Cone-Wesson & Ramirez, 1997; Small & Stapells, 2008a). A logical next step is to measure ABG differences in infants after applying infant-specific correction factors. This study reported ABR and ASSR threshold without the use of any correction factors as these are not yet available for the ASSR (especially for BC stimuli). Future research to establish ASSR eHL correction factors may assist in the clinical application of ASSR thresholds, and act as a springboard for future research involving ASSR ABG in infants with CHL.

Conclusion

EHDI Programs aim to identify hearing loss early in young infants, with many including mild hearing loss in their target population. Especially outside of North America, there continues to be the perception that obtaining ABR thresholds to low-frequency tone-bursts are too problematic for clinical use. This is not the experience within Canadian EHDI programs (Bagatto et al., 2020; Hatton et al., 2022), nor of the present study. Tone-evoked ABR for low- to high-frequency stimuli is indeed feasible and can be used to separate mild CHL from NH. In contrast, the ASSR may be more problematic. Compared to the ABR, the variability of ASSR thresholds and ABG size was too great to reliably separate NH from mild CHL, at least, using the parameters outlined. Furthermore, sensitivity and specificity for identifying CHL was highest for AC ABR threshold measurement compared to screening and ASSR threshold measurements. This finding supports continuation of the current practice using the ABR as the primary tool to assess hearing thresholds in young infants in Canadian EHDI programs. Before considering the ASSR as a diagnostic tool in this context, more research is needed using infants with varying degrees of CHL at multiple frequencies to fully assess its appropriateness.

References

Aimoni, C., Crema, L., Savini, S., Negossi, L., Rosignoli, M., Sacchetto, L., Bianchini, C., & Ciorba, A. (2018). Hearing threshold estimation by auditory steady state responses (ASSR) in children. *Acta Otorhinolaryngologica Italica, 38*(4), 361–368. https://doi.org/10.14639/0392-100X-1463

American National Standards Institute. (1996). *American national standard specifications for audiometers* (ANSI S3.6-1996).

Bagatto, M., Easwar, V., El-Naji, R., Hyde, M., Malandrino, A., Martin, V., Pigeon, M., Purcell, D., Scollie, S., & Witte, J. (2020). *Protocol for auditory brainstem response-based audiological assessment (ABRA)*. Ontario Infant Hearing Program. https://www.uwo.ca/nca/pdfs/clinical_protocols/2018.02 ABRA Protocol_Oct 2020.pdf

Casey, K.-A., & Small, S. A. (2014). Comparisons of auditory steady state response and behavioural air conduction and bone conduction thresholds for infants and adults with normal hearing. *Ear and Hearing, 35*(4), 423–439. https://doi.org/10.1097/AUD.0000000000000021

Cebulla, M., & Stürzebecher, E. (2015). Automated auditory response detection: Further improvement of the statistical test strategy by using progressive test steps of iteration. *International Journal of Audiology, 54*(8), 568–572. https://doi.org/10.3109/14992027.2015.1017659

Cone-Wesson, B., Parker, J., Swiderski, N., & Rickards, F. (2002). The auditory steady-state response: Full-term and premature neonates. *Journal of the American Academy of Audiology, 13*(5), 260–269. https://doi.org/10.1055/s-0040-1715969

Cone-Wesson, B., & Ramirez, G. M. (1997). Hearing sensitivity in newborns estimated from ABRs to bone-conducted sounds. *Journal of the American Academy of Audiology, 8*(5), 299–307. https://pubmed.ncbi.nlm.nih.gov/9328891/

Gorga, M. P., Johnson, T. A., Kaminski, J. R., Beauchaine, K. L., Garner, C. A., & Neely, S. T. (2006). Using a combination of click- and tone burst-evoked auditory brain stem response measurements to estimate pure-tone thresholds. *Ear and Hearing, 27*(1), 60–74. https://doi.org/10.1097/01.aud.0000194511.14740.9c

Hall, J. W. (2016). Crosscheck principle in pediatric audiology today: A 40-year perspective. *Journal of Audiology and Otology, 20*(2), 59–67. https://doi.org/10.7874/jao.2016.20.2.59

Hatton, J. L., Janssen, R. M., & Stapells, D. R. (2012). Auditory brainstem responses to bone-conducted brief tones in young children with conductive or sensorineural hearing loss. *International Journal of Otolaryngology, 2012,* Article 284864. https://doi.org/10.1155/2012/284864

Hatton, J. L., Rowlandson, J., Beers, A., & Small, S.A. (2019). Telehealth-enabled auditory brainstem response testing for infants living in rural communities: The British Columbia Early Hearing Program experience. *International Journal of Audiology, 58*(7), 381–392. https://doi.org/10.1080/14992027.2019.1584681

Hatton, J. L., & Stapells, D. R. (2011). The efficiency of the single- versus multiple-stimulus auditory steady state responses in infants. *Ear and Hearing, 32*(3), 349–357. https://doi.org/10.1097/AUD.0b013e3181ff352c

Hatton, J. L., Van Maanen, A., & Stapells, D. R. (2022). *British Columbia early hearing program: Auditory brainstem response protocol.* http://www.phsa.ca/bc-early-hearing/Documents/ABR_Protocol.pdf

Janssen, R. M., Usher, L., & Stapells, D. R. (2010). The British Columbia's Children's Hospital tone-evoked auditory brainstem response protocol: How long do infants sleep and how much information can be obtained in one appointment? *Ear and Hearing, 31*(5), 722–724. https://doi.org/10.1097/AUD.0b013e3181ddf5c0

Jerger, J. (1970). Clinical experience with impedance audiometry. *Archives of Otolaryngology, 92*, 311–324. https://doi.org/10.1001/archotol.1970.04310040005002

John, M. S., Brown, D. K., Muir, P. J., & Picton, T. W. (2004). Recording auditory steady-state responses in young infants. *Ear and Hearing, 25*(6), 539–553. https://doi.org/10.1097/01.AUD.0000148050.80749.AC

John, M. S., & Picton, T. W. (2000). Human auditory steady-state responses to amplitude-modulated tones: Phase and latency measurements. *Hearing Research, 141*(1–2), 57–79. https://doi.org/10.1016/S0378-5955(99)00209-9

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

Lins, O. G., Picton, T. W., Boucher, B. L., Durieux-Smith, A., Champagne, S. C., Moran, L. M., Perez-Abalo, M. C., Martin, V., & Savio, G. (1996). Frequency-specific audiometry using steady-state responses. *Ear and Hearing, 17*(2), 81–96. https://doi.org/10.1097/00003446-199604000-00001

Luts, H., Desloovere, C., & Wouters, J. (2006). Clinical application of dichotic multiple-stimulus auditory steady-state responses in high-risk newborns and young children. *Audiology & Neurotology, 11*(1), 24–37. https://doi.org/10.1159/000088852

Michel, F., & Jørgensen, K. F. (2017). Comparison of threshold estimation in infants with hearing loss or normal hearing using auditory steady-state response evoked by narrow band CE-chirps and auditory brainstem response evoked by tone pips. *International Journal of Audiology, 56*(2), 99–105. https://doi.org/10.1080/14992027.2016.1234719

Nagashima, H., Udaka, J., Chida, I., Shimada, A., Kondo, E., & Takeda, N. (2013). Air-bone gap estimated with multiple auditory steady-state response in young children with otitis media with effusion. *Auris Nasus Larynx, 40*(6), 534–538. https://doi.org/10.1016/j.anl.2013.03.002

Rance, G., & Rickards, F. (2002). Prediction of hearing threshold in infants using auditory steady-state evoked potentials. *Journal of the American Academy of Audiology, 13*(5), 236–245. https://doi.org/10.1055/s-0040-1715967

Rance, G., Roper, R., Symons, L., Moody, L.-J., Poulis, C., Dourlay, M., & Kelly, T. (2005). Hearing threshold estimation in infants using auditory steady-state responses. *Journal of the American Academy of Audiology, 16*(5), 291–300. https://doi.org/10.3766/jaaa.16.5.4

Rance, G., Tomlin, D., & Rickards, F. W. (2006). Comparison of auditory steady-state responses and tone-burst auditory brainstem responses in normal babies. *Ear and Hearing, 27*(6), 751–762. https://doi.org/10.1097/01.aud.0000240491.68218.ed

Rodrigues, G. R. I., & Lewis, D. R. (2014). Establishing auditory steady-state response thresholds to narrow band CE-chirps® in full-term neonates. *International Journal of Pediatric Otorhinolaryngology, 78*(2), 238–243. https://doi.org/10.1016/j.ijporl.2013.11.012

Rodrigues, G. R. I., Lewis, D. R., & Fichino, S. N. (2010). Steady-state auditory evoked responses in audiological diagnosis in children: A comparison with brainstem evoked auditory responses. *Brazilian Journal of Otorhinolaryngology, 76*(1), 96–101. https://doi.org/10.1590/s1808-86942010000100016

Sininger, Y. S., Hunter, L. L., Hayes, D., Roush, P. A., & Uhler, K. M. (2018). Evaluation of speed and accuracy of next-generation auditory steady state response and auditory brainstem response audiometry in children with normal hearing and hearing loss. *Ear and Hearing, 39*(6), 1207–1223. https://doi.org/10.1097/AUD.0000000000000580

Sininger, Y. S., Hunter, L. L., Roush, P. A., Windmill, S., Hayes, D., & Uhler, K. M. (2020). Protocol for rapid, accurate, electrophysiologic, auditory assessment of infants and toddlers. *Journal of the American Academy of Audiology, 31*(6), 455–468. https://doi.org/10.3766/jaaa.19046

Small, S. A., Hatton, J. L., & Stapells, D. R. (2007). Effects of bone oscillator coupling method, placement location, and occlusion on bone-conduction auditory steady-state responses in infants. *Ear and Hearing, 28*(1), 83–98. https://doi.org/10.1097/01.aud.0000249787.97957.5b

Small, S. A., & Stapells, D. R. (2008a). Maturation of bone conduction multiple auditory steady-state responses. *International Journal of Audiology, 47*(8), 476–488. https://doi.org/10.1080/14992020802055284

Small, S. A., & Stapells, D. R. (2008b). Normal ipsilateral/contralateral asymmetries in infant multiple auditory steady-state responses to air- and bone-conduction stimuli. *Ear and Hearing, 29*(2), 185–198. https://doi.org/10.1097/01.aud.0000305157.83012.f4

Stapells, D. R., & Ruben, R. J. (1989). Auditory brain stem responses to bone-conducted tones in infants. *Annals of Otology, Rhinology & Laryngology, 98*(12), 941–949. https://doi.org/10.1177/000348948909801205

Stapells, D. R., & Small, S. A. (2017). Threshold assessment in infants using the frequency-specific auditory brainstem response and auditory steady-state response. In A. M. Tharpe & R. C. Seewald (Eds.), *Comprehensive handbook of pediatric audiology* (2nd ed., pp. 505–550). Plural Publishing.

Swanepoel, D. W., Ebrahim, S., Friedland, P., Swanepoel, A., & Pottas, L. (2008). Auditory steady-state responses to bone conduction stimuli in children with hearing loss. *International Journal of Pediatric Otorhinolaryngology, 72*(12), 1861–1871. https://doi.org/10.1016/j.ijporl.2008.09.017

Swanepoel, D. W., & Steyn, K. (2005). Short report: Establishing normal hearing for infants with the auditory steady-state response. *The South African Journal of Communication Disorders, 52*, 36–39.

Van Maanen, A., & Stapells, D. R. (2009). Normal multiple auditory steady-state response thresholds to air-conducted stimuli in infants. *Journal of the American Academy of Audiology, 20*(3), 196–207. https://doi.org/10.3766/jaaa.20.3.6

Van Maanen, A., & Stapells, D. R. (2010). Multiple-ASSR thresholds in infants and young children with hearing loss. *Journal of the American Academy of Audiology, 21*(8), 535–545. https://doi.org/10.3766/jaaa.21.8.5

Widen, J. E., White, K. R., Gravel, J. S., Vohr, B. R., James, M., Kennalley, T., Maxon, A. B., Spivak, L., Sullivan-Mahoney, M., Weirather, Y., & Meyer, S. (2005). A multisite study to examine the efficacy of the otoacoustic emission/automated auditory brainstem response newborn hearing screening protocol: Results of visual reinforcement audiometry. *American Academy of Audiology, 14*, S2000–S2216. https://doi.org/10.1044/1059-0889(2005/022)

Yang, E. Y., Rupert, A. L., & Moushegian, G. (1987). A developmental study of bone conduction auditory brain stem response in infants. *Ear and Hearing, 8*(4), 244–251. https://doi.org/10.1097/00003446-198708000-00009

Yang, E. Y., Stuart, A., Stenstrom, R., & Green, W. B. (1993). Test-retest variability of the auditory brainstem response to bone-conducted clicks in newborn infants. *International Journal of Audiology, 32*(2), 89–94. https://doi.org/10.3109/00206099309071859

Yang, E. Y., Stuart, A., Stenstrom, R., & Hollett, S. (1991). Effect of vibrator to head coupling force on the auditory brain stem response to bone conducted clicks in newborn infants. *Ear and Hearing, 12*(1), 55–60. https://doi.org/10.1097/00003446-199102000-00007

Authors' Note

Correspondence concerning the article should be addressed to Hope Valeriote, Audiology Services, Glenrose Rehabilitation Hospital, 10230 111 Ave, Edmonton, AB, CANADA. **Email: hope.valeriote@ahs.ca**

Acknowledgments

This research was supported by an NSERC Discovery Grant to Dr. Small who sadly passed away in 2022. Thanks to Dr. Douglas for assistance obtaining ethics approval and supervision of data collection, Drs. Burkard and Stapells for mentorship in the preparation of this manuscript, and A. Gascon for assistance with data analysis.

Canadian Journal of Speech-Language Pathology and Audiology (CJSLPA)

Program Evaluation of a Low-Intensity Parent-Implemented Intervention for Young Children Late to Talk: How Much Is Enough?

Évaluation d'un programme d'accompagnement parental de faible intensité destiné à de jeunes enfants qui commencent à parler tardivement : quelle intensité est suffisante?

Allison Maier
Aamir Sholapur
Sarah Donnelly

KEYWORDS

PRESCHOOL SPEECH AND LANGUAGE

LATE TALKERS

PARENT-IMPLEMENTED INTERVENTIONS

FOCUS

Allison Maier[1], Aamir Sholapur[1,2], and Sarah Donnelly[1]

[1]Kingston, Frontenac, and Lennox & Addington Public Health, Kingston, ON, CANADA

[2]Department of Public Health Sciences, Queen's University, Kingston, ON, CANADA

Editor:
David H. McFarland

Abstract

There is a high prevalence of late talkers, many of whom catch up to their peers without intervention. As publicly funded programs often have long wait times, evidence is needed to inform intervention intensity for this population. The aim of the study was to evaluate a low-intensity (three session, 4 hr) parent-implemented intervention based on parent and child outcomes. Parents and children were recruited during the initial assessment for a pre–post quasiexperimental study. At initial and reassessment, parents (n = 67) completed a survey developed to measure confidence and behaviour and children's (n = 89) communicative participation outcomes were measured using the Focus on the Outcome of Communication Under Six. Participants were grouped based on different intervention intensity that resulted from family attendance: experimental (all three sessions), partial control group (some sessions), and full control group (no sessions). Paired t tests and analysis of variance were used to identify differences across time (pre–post) and group. In parents, paired t tests detected statistically significant increases in the experimental and partial control groups. Similarly, clinically and statistically significant differences in Focus on the Outcome of Communication Under Six scores were observed pre- and postintervention in the experimental and partial control groups in children. Analysis of variance revealed no significant differences between experimental, partial control, or full control groups. Although no treatment effect for the low-intensity parent-implemented intervention model was found, this study raises important considerations of future research needs and current program decision-making for clinicians, service providers, and researchers.

Abrégé

Il y a une forte prévalence d'enfants qui commencent à parler tardivement et plusieurs d'entre eux rattrapent leurs pairs sans qu'aucune intervention leur soit offerte. En raison des temps d'attente importants pour accéder aux programmes financés par le public, des données probantes sont nécessaires pour déterminer l'intensité des interventions devant être offertes à cette population. L'objectif de cette étude était d'évaluer l'effet du programme d'accompagnement parental de faible intensité (trois séances de quatre heures) sur les parents et les enfants. Les parents et les enfants inclus dans cette étude quasi expérimentale prétest-post-test ont été recrutés lors d'une évaluation initiale en orthophonie. Lors de cette évaluation initiale et lors de la réévaluation, les parents (n = 67) ont répondu à un questionnaire conçu pour mesurer leur assurance et leurs comportements. La participation communicative des enfants (n = 89) a été mesurée à l'aide de l'outil *Focus on the Outcome of Communication Under Six*. L'assiduité des participants aux séances du programme d'accompagnement parental a été utilisée pour les grouper en fonction de l'intensité de l'intervention qu'ils ont reçue : groupe expérimental (présence à l'ensemble des trois séances), groupe « partiellement » contrôle (présence à quelques séances) et groupe « entièrement » contrôle (présence à aucune séance). Des tests de Student pour échantillons appariés et des analyses de variance ont été utilisés pour analyser les différences dans le temps (pré-post) et entre les groupes. Chez les parents, les tests de Student pour échantillons appariés ont révélé une augmentation statistiquement significative dans le groupe expérimental et dans le groupe partiellement contrôle. De même, des différences cliniquement et statistiquement significatives ont été observées en ce qui concerne les scores du *Focus on the Outcome of Communication Under Six* complétés avant et après le programme d'accompagnement parental chez les enfants du groupe expérimental et ceux du groupe partiellement contrôle. L'analyse de la variance n'a révélé aucune différence significative entre le groupe expérimental, le groupe partiellement contrôle et le groupe entièrement contrôle. Bien qu'aucun effet n'ait été constaté pour le programme d'accompagnement parental à faible intensité, la présente étude soulève d'importantes questions sur les besoins futurs en matière de recherche, ainsi que pour les décisions que les cliniciens, les prestataires de services et les chercheurs doivent prendre actuellement.

Late talker is a term used to describe children aged 18–35 months with delayed expressive language skills, no known underlying cause, and typical skills in other areas of development (Rescorla, 2011; Singleton, 2018). The estimated prevalence of late talkers at 24 months is about 13% (Zubrick et al., 2007). Some of these children may also have a mild receptive language delay (Rescorla, 2011), slow word processing (LaTourrette et al., 2023), and other risk factors that may predict later poor language outcomes (Fisher, 2017; Morgan et al., 2020; Perry et al., 2022). Although many late talkers catch up with their peers by school age (Rice et al., 2008), some will present with a persistent language delay that is eventually diagnosed as a developmental language disorder (Bishop et al., 2016; Singleton, 2018). Because developmental language disorder is characterized by the presence of language difficulties significant enough to impact daily life and no association with other known causes or diagnoses (Bishop et al., 2016), early identification and early intervention for this population are critical.

One potential service delivery pathway for late talkers is parent-implemented intervention that involves teaching parents skills and strategies to use their role as language facilitators to support their children's language development. In general, evidence suggests that these intervention models are impactful. Several systematic reviews and meta-analyses reported significant gains in children's receptive and expressive language across several language constructs following parent-implemented intervention when compared to controls (DeVeney et al., 2017; Roberts & Kaiser, 2011, 2015; Roberts et al., 2019; Tosh et al., 2017). Further, these studies suggested that outcomes following parent-implemented interventions for young children presenting with language difficulties are in fact not significantly different than outcomes observed in clinician-implemented interventions (DeVeney et al., 2017; Roberts & Kaiser, 2011; Roberts et al., 2019; Tosh et al., 2017).

Despite the extent of the literature on parent-implemented interventions to support language development in late talkers, very little literature has explored appropriate intensity for these models (Tosh et al., 2017). Studies included in these reviews offered anywhere between 2 and 110 hr of parent training (Roberts & Kaiser, 2011; Roberts et al., 2019). Currently, the lowest intensity level of intervention typically studied in the literature for this target group is between five and seven parent sessions (Buschmann et al., 2015; Ciccone, 2012; Cunningham et al., 2019; Kruythoff-Broekman et al., 2019; Kwok et al., 2020). A recent study by Zulkifli et al. (2023) compared the treatment effect for a varying number of parent sessions and found intensity did not predict child language outcomes. Zulkifli et al. called for further research into the number of sessions necessary for a treatment effect. Additionally, findings from many of the studies investigating five to seven sessions were limited by lack of comparison groups to adequately control for maturation effects and natural language gains (Ciccone, 2012; Cunningham et al., 2019; Kwok et al., 2020; Zulkifli et al., 2023).

Lack of clarity on appropriate intensity and the efficacy of lower intensity models poses several challenges. Administering long, high-intensity parent-implemented models may be excessive for children who often catch up to their peers, recognizing that high-intensity intervention is not always better (Frizelle et al., 2021). In fact, lengthy parent-implemented models may place unnecessary time and resource strains on families, creating barriers to care and contributing to program attrition (Mytton et al., 2014). Provision of these longer format interventions may also face operational challenges due to resource constraints that limit speech-language pathologist and speech-language assistant time, and in turn, result in higher wait times for all services. Given these considerations, it is beneficial to investigate the efficacy of low-intensity (fewer than five sessions) parent-implemented interventions as a first intervention option for late-talking toddlers, given more intensive services could be offered afterwards, as needed, in response to the child's progress.

Addressing this research need, the current study evaluated a low-intensity parent-implemented intervention for late talkers in a publicly funded clinical setting in southeast Ontario. The goal of the program evaluation was to determine the most appropriate service delivery pathway in the clinic for late talkers by asking the following specific evaluation questions:

1. In a real-world clinical setting, what are the outcomes for children who are late talkers and parents who attend all, some, or none of a three-session parent-implemented intervention?
2. Were there differences in outcomes based on the level of attendance in the program?
3. What was the attendance rate for sessions and the overall attrition rate for the intervention?

Methods

Study Design

This evaluation used a pre–post design with a nonrandom quasicontrol group as, given the real-world clinic-based

setting, a randomized clinical trial design was not possible. The evaluation was designed to use the best available control group given the high likelihood of maturation in the target population (due to both their age and the possibility of them catching up to their age-matched peers without any direct intervention). Specifically, two convenience control groups were identified based on the number of intervention sessions attended after agreeing to participate: some (a partial control group) and none (a full control group).

The Queen's University Health Sciences and Affiliated Hospitals Research Ethics Board approved the research study (#6016667 KFLA-061-15). Informed consent was obtained from all participants at the time of recruitment using a standardized recruitment script, letter of information, and consent form.

Recruitment

The evaluation was conducted in a clinical preschool speech and language program setting funded by Ontario's Ministry of Children, Community and Social Services. All children in the program receive an initial assessment with a speech-language pathologist. The parents of children identified as late talkers between 18 and 30 months of age were recommended to attend the three-session parent-implemented intervention as the first intervention recommended in the service delivery pathway for this target group.

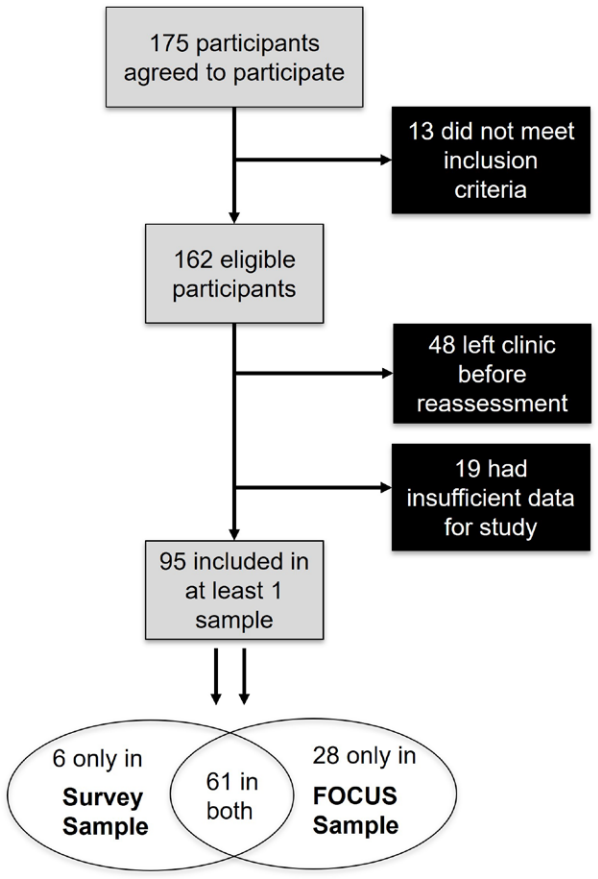
Speech-language pathologists were provided with the eligibility criteria, with late talkers defined as children with delayed expressive language in the absence of a causal diagnosis or delays noted in other areas of development, including social communication skills or receptive language skills. A notable exception was to include children who also have a mild receptive language delay identified through informal assessment. The program content and structure were designed to also meet the needs of this group of children and a more restrictive inclusion criteria would limit both treatment options for these families and enrolment in the treatment group. A formal assessment tool was not used, as this would have deviated from typical clinical practice for this age.

Starting in spring 2016, all intervention participants were invited to participate in the study. Families were included in the study if their child was 18 months at time of initial assessment and was not receiving additional interventions, including private speech-language therapy. Recruitment ceased in March 2020, before the intended sample size was met, due to changes in the intervention's service delivery model with the onset of the COVID-19 pandemic.

Participants

A total of 175 participants were recruited to the study, 13 of whom did not meet inclusion criteria. Of the remaining 162 eligible participants, 95 were included in the evaluation, 48 left the clinic before the end of the intervention, and 19 had insufficient data to be included. A full overview of the participants can be seen in **Figure 1**. Of the 95 final participants, the mean age was 107 ± 16 weeks (approximately 26 months ± 4 months); 52.6% of children included had a Communication Function Classification System score of Level III, 37.9% Level IV, 6.3% Level II, 2.1% Level V and 1.1% Level I. A more detailed breakdown of the participants as well as those who left the clinic and those who had insufficient data is shown in **Table 1**, including breakdowns by demographics, hearing status, and comorbidities. An observational analysis was conducted to compare those included in the study with those excluded to determine any differences. The children who left the clinic

Figure 1



Flowchart of recruited participants

Table 1

Summary of Included Participants, Those Who Left the Clinic Before the Intervention Ended, and Those Excluded Due to Insufficient Data

Characteristic	Included participants *N* = 95 % (*n*)	Left clinic before intervention ended *N* = 48 % (*n*)	Excluded due to insufficient data *N* = 19 % (*n*)
Age at initial assessment			
in weeks *M* ± *SD*	107 ± 16	99 ± 17	104 ± 19
Gender			
Male	72.6 (69)	64.6 (31)	57.9 (11)
Multilingual			
Yes	9.5 (9)	12.5 (6)	21.1 (4)
Communication Function Classification System level			
I	1.1 (1)	0	0
II	6.3 (6)	4.2 (2)	0
III	37.9 (36)	43.8 (21)	36.8 (7)
IV	52.6 (50)	33.3 (16)	47.4 (9)
V	2.1 (2)	2.1 (1)	0
Missing	0	16.7 (8)	15.8 (3)
Time spent in early learning environment			
None	60.0 (57)	52.1 (25)	47.4 (9)
0.5–2.5 days/week	6.3 (6)	8.3 (4)	0
≥ 2.5 days /week	33.7 (32)	22.9 (11)	36.8 (7)
Missing	0	16.7 (8)	15.8 (8)
Hearing			
No concerns	81.1 (77)	81.2 (39)	73.7 (14)
History of ear infections	2.1 (2)	2.1 (1)	15.8 (3)
Ongoing concerns	14.7 (14)	12.5 (6)	5.3 (1)
Missing	2.1 (2)	4.2 (2)	5.3 (1)
Comorbidities			
None	93.7 (89)	79.2 (38)	78.9 (15)
ASD	0	0	0
Anything confounding	4.2 (4)	2.1 (1)	0
Missing	2.1 (2)	18.8 (9)	21.1 (4)

Note. ASD = autism spectrum disorder.

before the end of the intervention were slightly younger than the other two populations. Otherwise, there were minor differences in the distribution of the groups, notably more multilingual children were excluded due to insufficient data.

For the 95 participants included, missing data remained an issue and so two samples were defined: a survey sample (67 participants) and a sample based on the Focus on the Outcomes of Communication Under Six

(FOCUS; Thomas-Stonell et al., 2012; 89 participants). A more detailed definition of the two samples is provided in the following sections. For both samples, the majority of participants were in the experimental group, followed in number by the partial control group, and then the control group. **Table 2** shows the breakdown of each sample and their characteristics overall and by group. Statistically it was unclear, but the two partial control groups may have had lower rates of hearing concerns.

Intervention Design

The intervention was originally developed in September 2014. The initial assessment, within the context of the Ontario Preschool Speech and Language Program, included specific requirements such as obtaining a relevant case history, assessing communication development, and in consultation with the family, recommending an appropriate intervention if the child demonstrated a communication delay. Suggestions and strategies for home were provided to parents as part of the initial assessment and, when appropriate and time allowed, were demonstrated with parent coaching to support understanding and learning. For late talkers, the intervention curriculum builds on the early language modelling strategies provided at the initial assessment.

The evaluated parent-implemented intervention consisted of three group sessions, typically over 3 to 5 months, for a total parent-training intervention time of 4 hr. The program design evolved from longer, more intensive parent-implemented interventions for this population. The number of hours and number of sessions were reduced from previous intervention designs due to parent attrition (attendance reduced over time), wait time management for young children (fewer sessions meant a new group could start every month), and general resource demands (managing wait times for all children).

Sessions were conducted by speech-language pathologists and speech-language assistants, and, in some cases, a community librarian or early literacy specialist provided Session 3. The content of the language modelling strategies included in Session 1 was described by Moharir et al. (2014). Handouts provided (e.g., "Face-to-Face") were consistent with the parent handouts referenced in Moharir et al.'s article and used with permission. Each session also identified personalized goals for specific language modelling strategies to be implemented during daily routines (Session 1 – adults only), while playing (Session 2 – adult and child coaching), and while reading books (Session 3 – adults only). **Table 3** contains a summary of intervention content. The strategies included in the sessions can be summarized into three categories:

- being face-to-face with their child when communicating,
- thinking about what they are saying and how they are saying it (e.g., simple language, repeating, using gestures, and imitating child's sounds/words), and
- watching and listening to how their child communicates (e.g., waiting, following child's lead, and paying close attention to child's words, sounds, and gestures).

The program theory supporting the intervention is outlined in **Figure 2**. Specifically, the intervention teaches parents how to implement the strategies alongside feedback to increase their confidence in doing so, with the aim of changing the parents' behaviours when interacting with their child during daily activities and routines. In turn, the children's outcomes are expected to improve. The evaluation of this intervention was designed according to the hypothesized program theory, especially in terms of parent outcomes.

After the intervention, children were reevaluated by a speech-language pathologist. Ideally, this occurred 6 months from the initial assessment to support timely re-administration of outcome measures and assess response to intervention to plan next intervention, as needed, but wait lists could increase the length between the two assessments.

Study Measures

The evaluation used both parent and child outcome measures which are summarized in **Table 4** alongside their collection points (typically the initial assessment for preintervention and reassessment for postintervention).

Parent Outcome Measures (Survey)

A self-reported parent outcome measure was used to assess the short- and intermediate-term intervention objectives. Using measures that focused on parent constructs that were not expected to change without intervention decreased the risk of maturation effects on the study. A survey tool was developed for this as the service delivery model did not include parent responsivity observations, given feasibility constraints in a clinic-based study. The survey was designed to be specific to the intervention, relate to the hypothesized program theory, be capable of detecting change pre- and postintervention, and use an appropriate literacy level.

The tool measured three constructs: the short-term outcome of confidence, the intermediate outcome of self-reported behaviour change, and the potential

Table 2

Summary of Survey and FOCUS Samples and Their Characteristics

Characteristic	Survey sample				FOCUS sample			
	Total included *N* = 67 % (*n*)	Experimental *N* = 42 % (*n*)	Control *N* = 11 % (*n*)	Partial control *N* = 14 % (*n*)	Total included *N* = 89 % (*n*)	Experimental *N* = 53 % (*n*)	Control *N* = 13 % (*n*)	Partial control *N* = 23 % (*n*)
Age at initial assessment								
in weeks *M* ± *SD*	105 ± 13	104 ± 11	103 ± 21	109 ± 11	107 ± 13	106 ± 11	103 ± 19	110 ± 13
Gender								
Male	70.1 (47)	73.8 (31)	63.6 (7)	64.3 (9)	73.0 (65)	77.4 (41)	61.5 (8)	69.6 (16)
Multilingual								
Yes	9.0 (6)	4.8 (2)	18.2 (2)	14.3 (2)	10.1 (9)	9.4 (5)	15.4 (2)	8.7 (2)
Time between initial & reassessment								
in weeks *M* ± *SD*	42 ± 15	41 ± 15	47 ± 19	44 ± 16	43 ± 17	41 ± 16	48 ± 16	43 ± 16
Missing or N/A	4.4 (3)	4.7 (2)	9.1 (1)	0	2.2 (2)	1.9 (1)	7.7 (1)	0
Communication Function Classification System level								
I	1.5 (1)	2.4 (1)	0	0	0	0	0	0
II	7.5 (5)	11.9 (5)	0	0	6.7 (6)	9.4 (5)	7.7 (1)	0
III	29.9 (20)	23.8 (10)	36.4 (4)	42.9 (6)	39.3 (35)	35.8 (19)	38.5 (5)	47.8 (11)
IV	59.7 (40)	61.9 (26)	54.5 (6)	57.1 (8)	52.8 (47)	52.8 (28)	53.8 (7)	52.2 (12)
V	1.5 (1)	0	9.1 (1)	0	1.1 (1)	1.9 (1)	0	0
Time spent in early learning environment								
None	58.2 (39)	54.8 (23)	54.5 (6)	71.4 (10)	61.8 (55)	60.4 (32)	46.2 (6)	73.9 (17)
0.5–2.5 days/week	7.5 (5)	7.1 (3)	9.1 (1)	7.1 (1)	4.5 (4)	3.8 (2)	7.7 (1)	4.3 (1)
≥ 2.5 days /week	34.3 (23)	38.1 (16)	36.4 (4)	21.4 (3)	33.7 (30)	35.8 (19)	46.2 (6)	21.7 (5)

Table 2 (continued)

Summary of Survey and FOCUS Samples and Their Characteristics								
Hearing								
No concerns	77.6 (52)	71.4 (30)	72.7 (8)	100 (14)	80.9 (72)	75.5 (40)	69.2 (9)	100 (23)
History of ear infections	1.5 (1)	2.4 (1)	0	0	2.2 (2)	3.8 (2)	0	0
Ongoing concerns	19.4 (13)	23.8 (10)	27.3 (3)	0	14.6 (13)	18.9 (10)	23.1 (3)	0
Missing	1.5 (1)	2.4 (1)	0	0	2.2 (2)	1.9 (1)	7.7 (1)	0
Comorbidities								
None	92.5 (62)	92.9 (39)	90.9 (10)	92.9 (13)	95.5 (85)	96.2 (51)	92.3 (12)	95.7 (22)
ASD	0	0	0	0	0	0	0	0
Anything confounding	4.5 (3)	2.4 (1)	9.1 (1)	7.1 (1)	4.5 (4)	3.8 (2)	7.7 (1)	4.3 (1)
Missing	3.0 (2)	4.8 (2)	0	0	0	0	0	0
Parent relationship [a]								
Father/stepfather	10.4 (7)	16.7 (7)	0	0	6.7 (6)	11.3 (6)	0	0
Mother/stepmother	85.1 (57)	81.0 (34)	90.9 (10)	92.9 (13)	61.8 (55)	60.4 (32)	76.9 (10)	56.5 (13)
Missing	4.5 (3)	2.4 (1)	9.1 (1)	7.1 (1)	31.5 (28)	28.3 (15)	23.1 (3)	43.5 (10)
Parent education								
Postsecondary diploma/degree	76.1 (51)	76.2 (32)	81.8 (9)	71.4 (10)	53.9 (48)	56.6 (30)	61.5 (8)	43.5 (10)
Less than postsecondary degree & other	19.4 (13)	21.4 (9)	9.1 (1)	21.4 (3)	14.6 (13)	15.1 (8)	15.4 (2)	13.0 (3)
Missing	4.5 (3)	2.4 (1)	9.1 (1)	7.1 (1)	31.5 (28)	28.3 (15)	23.1 (3)	43.5 (10)

Note. ASD = autism spectrum disorder; FOCUS = Focus on Outcomes of Communication Under Six (Thomas-Stonell et al., 2012).
[a] Response categories that had no selection: grandparent, other relative, guardian/foster parent, and other.

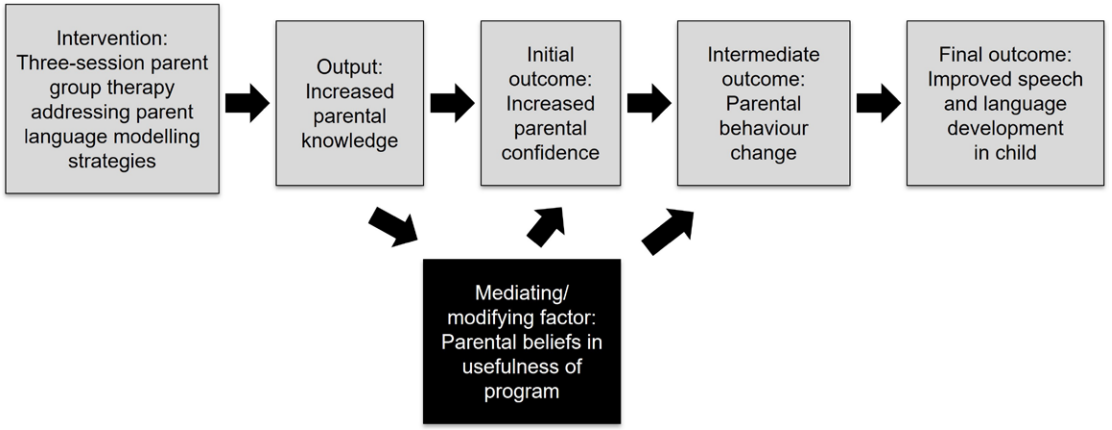
mediating factor of belief in the intervention. These constructs are similar to measures used in "readiness rulers" that have been applied in a wide variety of health-related lifestyle behaviour change settings (Bethelot et al., 2012; Rollnick et al., 2009). The two primary constructs (confidence and self-reported behaviour) were asked for each of the three main strategies employed by the intervention (being face-to-face, thinking about what you say and how you say it, and watching and listening: e.g., How confident are you in your ability to be face-to-face with your child when communicating? How often do you think about what you are saying and how you are saying it?). A final question asked the level of belief in the intervention (How much do you believe that by you using communication strategies with your child, you will improve your child's language skills?). Each question was asked on a 10-point Likert-style scale (1 = *Rarely when I interact with my child* to 10 = *Every time I interact with my child*). Face validity and acceptability of the tool were sought through brief cognitive interviewing

Table 3

Summary of the Parent Training Intervention

Aspect of training	Initial assessment (up to 1.5 hr)	Session 1 (2 hr)	Session 2 (1 hr)	Session 3 (1 hr)	Reassessment (up to 1.5 hr)
Participants	Child and parent(s)/ guardian(s).	Parents/guardians.	Children and parents/ guardians.	Parents/guardians.	Child and parent(s)/guardian(s).
Goal setting for parent/ guardian	Speech-language pathologist and parent work together to identify one or more goal.	Parent/guardian identifies one specific strategy that they will implement during daily routines with their child.	Parent/guardian identifies one specific strategy that they will implement while playing with their child.	Parent/guardian identifies one specific strategy that they will implement while reading books with their child.	Individualized.
Goal setting for child	Individualized.	With help from speech-language pathologist, parent chooses target vocabulary or 2 different 2-word combination types (e.g., descriptive word + noun).	Review goals set in Session 1 and discuss progress. Target goals in session through play.	Parents/guardian selects new target words or 2-word combinations to target with selected book.	Individualized.
Language modelling strategies targeted to parent/ guardian	Identify up to 3 strategies provided (with handouts).	Teach strategies and discuss use within routines (e.g., face-to-face).	Review strategies and coach use within play activities. May model and coach some new strategies in play, as needed.	Review strategies and discuss use within reading activities.	Review strategies with parents. May introduce new strategies at this time to address new goals beyond group.
Techniques used by facilitator	Play-based using formal and informal tools. Discussion of results with parent/guardian. Modelling and discussion of language modelling strategies.	Interactive discussions in small group setting. Hands-on practice. Handouts provided for each strategy.	Coach parent(s) during facilitated play with child. Play-based activities to practice language modelling strategies.	Interactive discussions in small group setting. Hands-on practice. Handout booklet provided. Free take-home book.	Play-based using formal and informal tools. Speech-language pathologist reviews and documents progress of child's goals from group. Discussion of results with parent/ guardian. Identify new child goals and new intervention plan, based on progress.

Figure 2



Program theory of low-intensity intervention

Table 4

Summary of Study Measures

Instrument	Description	Completed by	Completed on	Data-collection point for premeasure	Data collection point for postmeasure
Focus on the Outcomes of Communication Under Six (FOCUS) Parent Form (Washington, 2015)	Participation-based child measure that links speech and language treatment to the child's ability to communicate and participate in their world	Parent	Child	Initial assessment	Reassessment
MacArthur-Bates Communicative Development Inventories (Fenson, 2006)	Impairment-based child measure that screens language and communication skills	Parent	Child	1st intervention session	Reassessment
Reassessment recommendation	Postintervention outcome recommended for child (discharge or type of future intervention)	Speech-language pathologist	Child	N/A	Reassessment
Parent survey	Measures parental confidence, behaviour, and beliefs in intervention	Parent	Parent	End of initial assessment [a]	Reassessment

[a] To gather complete data, some parents completed the survey at home after the initial assessment or in the waiting room before the first intervention session.

with families who were participating in the intervention during the study design phase. These interviews continued until saturation was reached (six families) and the number and wording of questions was finalized to those described above based on this parent feedback.

Child Outcome Measures (FOCUS)

The FOCUS (Thomas-Stonell et al., 2012) was used as the child-based outcome measure, a validated tool that measures the child's overall communicative participation (Washington et al., 2015). During the study, the tool changed from being 50 items to 34 items (Oddson et al., 2019). The study used the standardized methods provided to convert between the two (Thomas-Stonell et al., 2012). A minimal clinically important difference (MCID) of 11 points for the 34-item tool has been established (Oddson et al., 2019; Washington et al., 2015). At time of study development, evidence suggested that the tool was robust against maturation effects, especially at the level of the MCID (Thomas-Stonell, 2013; Washington et al., 2015), an important consideration given the study used a pre–post design.

The evaluation also intended to measure change in the child's expressive vocabulary using the MacArthur-Bates Communication Development Inventory (Fenson, 2006) as an impairment-based measure, but due to high amount of missing data for all groups, these data were not analyzable. Instead, a brief supplementary analysis was performed looking at next intervention recommendations at reassessment (whether the child was discharged/recommended for home monitoring, recommended for an additional group therapy, or recommended for individual therapy). The type and intensity of intervention recommended at reassessment was informed primarily by child outcomes of the previous intervention but may also have been informed by other child and family factors (e.g., family schedule).

Outcome Analyses

Participants were included in the outcome analysis if there were sufficient data to assess the inclusion criteria, their attendance at the intervention, and pre–post data for at least one of the child or parent outcome measures. To maximize sample sizes, two sample sets were defined for analyses: the survey sample (having pre–post parent outcome measures) and the FOCUS sample (having pre–post child outcome measures). Participants were also classified into one of three groups: experimental (attended all three sessions), full control (attended none of the sessions despite agreeing to attend), and partial control (attended some but not all three sessions). **Figure 1** and **Table 2** list the numbers in each group. All analyses were performed in R (version 4.1.1).

One primary outcome measure was defined for the parent survey – a total change in score across all confidence and self-reported behaviour questions. Supplementary exploratory measures were defined for a change in each of the constructs (a change in score across all confidence questions and a change in score across all self-reported behaviour questions) and for change in score on the strategies (being face-to-face, thinking about what you say and how you say it, and watching and listening). The final survey question, the potential mediating factor of beliefs, was not used as there was insufficient sample size to conduct mediating factor analyses. A single outcome measure was defined for children – the mean difference in FOCUS scores.

To answer the first evaluation question (the outcomes in parents and children depending on number of sessions attended), individual paired *t* tests were conducted for each outcome measure for each group with the mean differences, 95% confidence intervals and *p* values reported. To control for Type 1 error, the *p* values were interpreted in the context of the number of analyses conducted. For example, the parent analyses were considered at a *p*-value cutoff of .003 (.05 divided by the 18 total analyses). For the child measure, a significant increase of the MCID signified a likely improvement. For the parent measures, any statistically significant increase signified a likely improvement. To answer the second evaluation question (differences based on number of sessions attended), analysis of variance (ANOVA) tests were used to determine if there were any differences between experiment, full, and partial control groups. For the parent measure, the ANOVA was only used on the primary outcome measure. To further investigate question two, χ^2 analyses were used to assess differences by groups for the reassessment recommendations.

Intervention Completion Analyses

To assess the final evaluation question (the intervention attendance and attrition rate), a final sample was identified that contained all participants with complete attendance records. From this, the total completion rate for the intervention (the percentage of participants who attended all sessions) and the overall session absence rate (the percentage of all sessions that were not attended) were determined.

Furthermore, for all children who did not complete the intervention or reassessment, information was retrieved

from the child's record to determine any known reason for the absence (e.g., participant moved to a new region or the clinic was unable to reach the participant's family). This information was used to provide context to the high rate of individuals excluded from the outcome analyses as well as for understanding the implications to delivering multisession group therapy in the clinical setting.

Results

Parent Outcome

The parent survey results of the paired *t* tests analyses for the primary and supplementary survey outcome measures (related to confidence and self-reported behaviour) are found in **Table 5**. For the experimental and partial control groups, statistically significant mean differences between pre- and postsurvey ratings were observed for the primary outcome measure (total score). For the experimental group, there were also statistically significant differences in the supplementary measures for confidence, think about what you say and how you say it, and watch and listen. For the control group, statistically significant differences were not observed for any measures using the Type 1 error-adjusted *p* value. Observationally, the mean differences were similar between the experimental and control groups, but there was a smaller sample size for the control group. A post hoc analysis revealed, given the group's mean difference and standard deviation, that a sample size of 21 would have been required (compared to the actual sample size of 11). Both observationally and statistically (through ANOVA tests), there were no statistically significant differences in the means between the groups ($F(2,64) = 1.03$, $p = .362$).

Child Outcome

Table 6 shows the FOCUS score mean differences by FOCUS sample groups. All estimates of mean difference are the equivalent of at least 3 times the MCID of 11. Statistically significant differences above the MCID were observed for both the experimental and partial control groups. For the full control group, statistically significant differences were not observed, but a post hoc analysis revealed that a sample size of 26 would have been required (compared to the actual sample size of 13) given the group's mean difference and standard deviation. There were no differences between the groups either observationally or through ANOVA statistical testing ($F(2,86) = 1.83$, $p = .166$). There were also no differences by group for reassessment recommendations based on χ^2 analysis (discharge/home monitoring, additional group therapy, or individual therapy; $\chi^2(4, N = 11) = 2.16$, $p = .71$).

Intervention Attendance and Attrition

Of the 48 participants who left the clinic before the reassessment,

- 45.8% (22) were unreachable by the clinic (by phone or mail; at least two attempts were made per clinic policy)
- 22.9% (11) left due to parent reporting no further concerns or needs
- 18.8% (9) moved or were transferred to a different clinic
- 12.5% (6) declined a reassessment

Complete attendance data were available for 144 participants. Based on this, the intervention had a completion rate of 52.1%, with 20.8% not attending any of the intervention despite agreeing to do so at the initial assessment. Overall, 33.1% of all sessions were not attended (143 missed sessions out of 432 total sessions). As part of service delivery, significant efforts were made to ensure attendance, including reminder calls before Sessions 1 and 3 (Session 2 was scheduled shortly after session one so a reminder call was not made). Families who missed sessions were offered a spot in the subsequent group session(s), which were usually provided the following month.

Discussion

This program evaluation assessed the impact of a low-intensity parent-implemented intervention model in late talkers and their parents in the context of determining the most appropriate service delivery pathway for these children in a publicly funded clinic. It was the first to investigate a low-intensity model with nonrandom control groups. Overall, the study did not detect a treatment effect that could be solely associated with the intervention in either parents or children. Instead, statistically significant improvements of at least 3 times the validated MCID were observed in children that attended at least one session of the intervention (experimental group and partial control group). Furthermore, although the study lacked the power to detect statistical significance in the group who did not attend the intervention, observational results suggested a similar rate of improvement in this control group. The parent outcome results aligned with the child outcome findings. The reassessment recommendations similarly found no differences between groups. Meanwhile, high rates of attrition were observed throughout the program.

Unexpectedly, it is possible that all of the low-intensity intervention groups saw similarly improved outcomes.

Table 5

Parent Outcome Measures Paired *t* Test Mean Differences and 95% Confidence Intervals with Statistical Tests and Significance

Measure	Experimental group pre–post survey *M* [95% CI] *t* (degrees of freedom) *p* value	Control group pre–post survey *M* [95% CI] *t* (degrees of freedom) *p* value	Partial control group pre–post survey *M* [95% CI] *t* (degrees of freedom) *p* value
Total score	2.8 [1.4, 4.2]* 4.0 (41) < .001	3 [−1.5, 7.5] 1.5 (10) .171	5.1 [1.5, 8.8]* 3.02 (13) .001
Construct			
Confidence	1.2 [0.5, 1.9]* 3.4 (41) .002	1.2 [−1.2, 3.6] 1.1 (10) .295	2.4 [0.3, 4.5] 2.5 (13) .027
Self-reported behaviour	1.6 [0.6, 2.7] 3.2 (41) .003	1.8 [−0.7, 4.4] 1.6 (10) .148	2.7 [0.9, 4.5] 3.2 (13) .007
Strategy			
Face-to-face	−0.1 [−0.7, 0.5] −0.2 (41) .816	−0.5 [−2.9, 2.0] −0.4 (10) .690	0.9 [−0.4, 2.3] 1.5 (13) .166
Think about what you say and how you say it	1.7 [1.0, 2.4]* 4.9 (41) < .001	1.7 [0.3, 3.2] 2.6 (10) .026	2.7 [1.0, 4.6] 3.3 (13) .006
Watch and listen	1.2 [0.5, 1.9]* 3.6 (41) < .001	1.7 [0.1, 3.3] 2.4 (10) .036	1.4 [−0.1, 3.0] 2.0 (13) .068

* To adjust for Type 1 error, $p < 0.003$ was considered significant.

Table 6

FOCUS Score Mean Differences From Paired t Test with 95% Confidence Intervals and Statistical Significance

Group	*M* [95% CI]	*t* (degrees of freedom)	*p*
Experimental	45.2 [32.3, 58.1]*	7.1 (52)	< .001
Control	38.3 [−1.9, 78.6]	2.1 (12)	.060
Partial control	66.9 [44.7, 89.1]*	6.3 (22)	< .001

Note. FOCUS = Focus on Outcomes of Communication Under Six (Thomas-Stonell et al., 2012).
* To adjust for Type 1 error, $p < .017$ was considered significant.

There are three possible explanations for these findings: (a) that the evaluation measured a considerable maturation effect in all groups; (b) the control groups were exposed to a sufficient intervention to have a clinical improvement from it; or (c) a combination of both (a) and (b) occurred. Each postulation and its relative likelihood

are assessed below based on the detailed findings of this evaluation and recent literature.

In general, maturation effects in the study's child population are highly probable given the rapid speech and language development in typically developing children between ages 18 and 30 months (Ellis & Thal, 2008). At the time of study design, the FOCUS was reported to be largely immune to maturation effects and that a difference above the MCID was only due to an intervention (Thomas-Stonell et al., 2013; Washington et al., 2015). Based on this rationale, the FOCUS was selected as the outcome measure for this study. More recent literature does suggest a maturation effect is possible, at least in certain populations (Cunningham et al., 2019). Additionally, the FOCUS was validated with assessment every 6 months (Thomas-Stonell et al., 2012). Due to the realities of resource constraints in clinical settings, the mean length of time before pre- and post-FOCUS was 43 weeks (10.5 months). This extended time frame increases the probability of a maturation effect occurring. As such, it is reasonable to conclude that at least some of the improvements observed in children from all groups were due to a maturation effect.

There are, however, two reasons why maturation is unlikely to account for the entirety of the improvements observed by this study. First, improvements equivalent to multiple times the MCID were observed and so a maturation effect, even with the extended timeline, seems unlikely to explain the whole effect size. Second, there is no reason to expect a maturation effect in parents and so the improvements observed in the parent outcome measure cannot be explained by this phenomenon. This was one of the primary reasons for including the parent survey in the study.

Before the evaluated intervention and alongside the initial point of data collection, participants in all groups received an initial assessment. As described above, the initial assessment had elements that could be considered a parent-implemented intervention, including provision of early language modelling strategies along with demonstration and coaching, as time allowed in the session. It could be argued that the evaluated intervention was an extension and reiteration of the strategies (e.g., watch and listen) first introduced during the initial assessment. Hence, it is possible that the improvements detected by this study in both the parents and children were due to the initial assessment acting as an intervention in and of itself. As no data were collected before the initial assessment or after the initial assessment and before the first session, it is impossible to directly assess any treatment effect of the initial assessment in isolation from the three-session studied intervention. As the only statistical evidence of similar clinical improvements was between the partial and experimental groups, it is also possible attending some of the evaluated intervention sessions were equivalent to all sessions. In the recent study that identified a possible maturation effect in the FOCUS, a clinical improvement was observed between the initial assessment and intervention start suggesting this could also have been measuring an improvement due to the initial assessment (Cunningham et al., 2019).

Based on the above evidence, it is the researchers' postulation that the study findings are indicative of both a maturation effect in the target population and that the initial assessment and/or partial attendance to the intervention may be a sufficient intervention to achieve a clinical improvement in children and parents. As the primary objective of this evaluation was to determine the appropriate service delivery pathway for late talkers, this would suggest that interventions with fewer than three sessions may be acceptable.

The findings of this study raise an additional consideration. The partial and full control groups were critical to correctly interpreting the results. Without partial and full control groups to compare to the experimental group outcomes, statistically significant gains in both parent and child experiment groups could have been wrongly attributed to the intervention. Unfortunately, many of the higher intensity models in the literature for this target population use similar child outcome measures without the benefit of a control group yet make conclusions about treatment effects (Cunningham et al., 2019; Gaines & Gaboury, 2004; Kwok et al., 2020). This study reinforces the importance of study design, demonstrating that it is inappropriate to draw conclusions about treatment effects without a control group. The need for further research utilizing a control group was recently highlighted by one of these studies (Cunningham et al., 2019).

The overall goal of the evaluation was to determine the most appropriate service delivery pathway for late talkers. The attendance and attrition results provide critical context. Overall, only 52% of participants completed the intervention with one third of all sessions being missed. Hence, even a three-session intervention may be too many sessions to expect families to attend, for any number of reasons. Interventions with a higher number of sessions would likely have an even higher rate of attrition. This finding is supported in the literature; for example, Mytton et al. (2014) observed that competing demands were a significant barrier to parent-implemented models, including childcare and the frequency and timing of sessions in relation to the parents' work patterns. Hence, the

clinic-level efficiency of offering multisession group therapy when attrition is so high must be considered. To design the best service delivery pathway for this target population, a resource balance must be achieved between the additional benefit of each session with the additional resources required to offer it, accounting for expected attrition.

In the real-world clinical setting, such as the one in which the evaluation was conducted, the findings could be implemented to change the first intervention in one of two ways: (a) decrease the content of (and resources allocated to) the initial assessment so it ceases acting as an intervention and immediately refer children to the intervention or (b) focus efforts on the initial assessment as the intervention. The first choice could decrease wait times for the initial assessment, but with the high rate of attrition, could result in insufficient intervention intensity and an inefficient use of resources. To successfully implement the first choice, options to increase attendance of the parent-implemented intervention should be considered. Reduced resources allocated to the initial assessment (due to less content, shorter in duration) could be reinvested to reduce wait time for services. The second choice could also decrease wait times by decreasing the number of clinic staff required to provide the intervention and reallocate this time to conduct initial assessments and other services. Based on the findings of this program evaluation, a hybrid of these two options was implemented to change the service delivery pathway for late talkers in this clinic. The initial assessment format was kept intact by providing parents with links to online, recorded training modules for Sessions 1 and 3, prepared and maintained by a speech-language pathologist. These modules are made available to the parents immediately following the initial assessment (no wait time). If the parent needs further support to implement the strategies, one individual coaching session is offered as follow-up to the initial assessment intervention and online training modules. A reassessment to monitor progress and determine the next appropriate type and intensity of intervention continues to be scheduled for all children.

Limitations

The real-world clinical setting of the evaluation was both a strength and limitation. The practicalities of the clinical setting limited the rigour of the study both in design (a convenience sample, nonrandom allocation into groups, self-reported outcomes measures) and in implementation (limited sample sizes and group distribution, and an increase in time between data collection points). Attempts were made to minimize the impact of these limitations, including investigating differences between groups. There were only minor variations between the groups, but it is possible that the groups differed in nonmeasured ways. Because the parent outcome measure was a locally developed purpose-built tool, it did not undergo comprehensive psychometric testing. It was deemed to be fit for purpose as it provided additional support for findings from the FOCUS, which is fully validated. However, the clinical setting allowed for the broader program evaluation lens for the study, including studying attrition, which increased the applicability of the findings.

The early end to recruitment caused by the pandemic resulted in an insufficient sample size, especially in the control group, to analyze the evaluation data as intended in many instances. As evidenced by the post hoc analyses, the study did not have sufficient power to detect statistical changes in the full control group for either the parent or child outcome measures. As such, all results from this group were solely observational and do not provide the desired level of evidence to inform service delivery planning. However, the inclusion of this group in conjunction with the partial control group provides critical contextual information on the importance of a control group when studying, as discussed previously.

Furthermore, the evaluation was initially designed to control for potential confounding and/or mediating demographic and individual risk factor variables between the groups and the outcome. Unfortunately, given the final sample sizes, this was not possible. The researchers postulate that parent and/or child characteristics exist which could increase/decrease the likelihood of success of the intervention as evidenced in recent literature. In a study on children with language delays, Zulkifli et al. (2023) found that children and parents with a higher baseline had better outcomes. The study also noted that although measured parent responsivity was lower in multilingual parents, the intervention had been designed for the cultural norms of English-speaking families. Zulkifli et al. concluded that adapting interventions based on baseline levels could be possible, and that interventions should be better adapted to the linguistic and cultural needs of families. It is possible that subgroups of the population included in this evaluation would benefit from an adapted intervention resulting in more clinical effects than were observed in the overall group in this study. More research is required, especially that which can account for the feasibility of offering these adaptations in a real-world clinical setting.

Recommendations for Future Research and Clinical Practice

This evaluation observed a clinical improvement in children and parents who attended at least one session

of a three-session parent-implemented intervention, but was unable to measure a specific treatment effect from the intervention itself because control and partial control groups for children and parents showed similar gains. In consideration of the high attrition rates, the findings from this study were used to change the local service delivery pathway for late talkers. Similarly, these results can be used by clinicians and service delivery decision-makers to help determine appropriate pathways in their own service systems. Notably, these groups should reconsider the current best practice of higher intensity parent-implemented intervention models as a default recommendation because this evaluation both calls into question the evidence from these studies (with their lack of control groups) and provides early evidence in support of lower intensity first-intervention models for late talkers. Furthermore, decision-makers especially should consider the high rates of attrition in programming when determining efficient pathways and as such, consider few-to-no-session interventions while monitoring progress over time for this population.

Although the findings of this study are informative in clinical settings, they are limited in scope and by methodology. Further research is required on all possible intervention intensities, from initial assessments with strategies and coaching (as described in this study) to low-intensity models to higher intensity models. All research in this target population must include a control group to be able to establish a treatment effect, as the outcome measures are susceptible to maturation effects. However, wherever possible, real-world clinic settings should be used for this research as this evaluation demonstrates it is critical to understanding the feasibility of such interventions.

Conclusions

This study was the first to evaluate a low-intensity (three-session) parent-implemented intervention for late talkers in a publicly funded clinical setting, with control groups and in consideration of varying levels of attendance. Outcomes improved for both children and parents who attended some or all the sessions, with no statistically significant differences between groups, and observational evidence suggested similar outcomes for families who attended none of the sessions. The program had substantial attrition, further calling into question the feasibility of the service delivery pathway. The results from this program evaluation were used to change the service delivery pathway for late talkers and can inform similar changes in other service systems. The study demonstrated the need for further clinic-based research, where findings can simultaneously assess clinical outcomes and feasibility of intervention modalities.

References

Bertholet, N., Gaume, J., Faouzi, M., Gmel, G., & Daeppen, J. B. (2012). Predictive value of readiness, importance, and confidence in ability to change drinking and smoking. *BMC Public Health, 12,* Article 708. https://doi.org/10.1186/1471-2458-12-708

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE Consortium. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLOS ONE, 11*(7), Article e0158753. https://doi.org/10.1371/journal.pone.0158753

Buschmann, A., Multhauf, B., Hasselhorn, M., & Pietz, J. (2015). Long-term effects of a parent-based language intervention on language outcomes and working memory for late-talking toddlers. *Journal of Early Intervention, 37*(3), 175–189. https://doi.org/10.1177/1053815115609384

Ciccone, N., Hennessey, N., & Stokes, S. F. (2012). Community-based early intervention for language delay: A preliminary investigation. *International Journal of Language & Communication Disorders, 47*(4), 467–470. https://doi.org/10.1111/j.1460-6984.2012.00149.x

Cunningham, B. J., Kwok, E., Earle, C., & Oram Cardy, J. (2019). Exploring participation and impairment-based outcomes for Target Word: A parent-implemented intervention for preschoolers identified as late-to-talk. *Child Language Teaching and Therapy, 35*(2), 145–164. https://doi.org/10.1177/0265659019846931

DeVeney, S. L., Hagaman, J. L., & Bjornsen, A. L. (2017). Parent-implemented versus clinician-directed interventions for late-talking toddlers: A systematic review of the literature. *Communication Disorders Quarterly, 39*(1), 293–302. https://doi.org/10.1177%2F1525740117705116

Ellis, E. M., & Thal, D. J. (2008). Early language delay and risk for language impairment. *Perspectives on Learning and Education, 15*(3), 93–100. https://doi.org/10.1044/lle15.3.93

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2006). *MacArthur-Bates Communication Development Inventory* (2nd ed.). Brookes Publishing.

Fisher, E. L. (2017). A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *Journal of Speech, Language, and Hearing Research, 60*(10), 2935–2948. https://doi.org/10.1044/2017_JSLHR-L-16-0310

Frizelle, P., Tolonen, A.-K, Tulip, J., Murphy, C.-A., Saldana, D., & McKean, C. (2021). The influence of quantitative intervention dosage on oral language outcomes for children with developmental language disorder: A systematic review and narrative synthesis. *Language, Speech, and Hearing Services in Schools, 52*(2), 738–754. https://doi.org/10.1044/2020_lshss-20-00058

Gaines, B. R., & Gaboury, I. (2004). Toddler Talk: Outcomes from a parent-focused intervention for children with speech/language problems. *Canadian Journal of Speech-Language Pathology and Audiology, 28*(4), 173–183. https://www.cjslpa.ca/files/2004_JSLPA_Vol_28/No_04_151-194/Gaines_Gaboury_JSLPA_2004.pdf

Kruythoff-Broekman, A., Wiefferink, C., Rieffe, C., & Uilenburg, N. (2019). Parent-implemented early language intervention programme for late talkers: Parental communicative behaviour change and child language outcomes at 3 and 4 years of age. *International Journal of Language & Communication Disorders, 54*(3), 451–464. https://doi.org/10.1111/1460-6984.12451

Kwok, E., Cunningham, B. J., & Oram Cardy, J. (2020). Effectiveness of a parent-implemented language intervention for late-to-talk children: A real-world retrospective clinical chart review. *International Journal of Speech-Language Pathology, 22*(1), 48–58. https://pubmed.ncbi.nlm.nih.gov/30897967

LaTourrette, A., Waxman, S., Wakschlag, L. S., Norton, E. S., & Weisleder, A. (2023). From recognizing known words to learning new ones: Comparing online speech processing in typically developing and late-talking 2-year olds. *Journal of Speech, Language, and Hearing Research, 66*, 1658–1677. https://doi.org/10.1044/2023_JSLHR-22-00580

Moharir, M., Barnett, N., Taras, J., Cole, M., Ford-Jones, E. L., Levin, L., Barnett, N., & Taras, J. (2014). Speech and language support: How physicians can identify and treat speech and language delays in the office setting. *Paediatrics & Child Health, 19*(1), 13–18. https://doi.org/10.1093/pch/19.1.13

Morgan, L., Delehanty, A., Dillon, J. C., Schatschneider, C., & Wetherby, A. M. (2020). Measures of early social communication and vocabulary production to predict language outcomes at two and three years in late-talking toddlers. *Early Childhood Research Quarterly, 51,* 366–378. https://doi.org/10.1016/j.ecresq.2019.12.005

Mytton, J., Ingram, J., Manns, S., & Thomas, J. (2014). Facilitators and barriers to engagement in parenting programs: A qualitative systematic review. *Health Education & Behavior, 41*(2), 127–137. https://doi.org/10.1177/1090198113485755

Oddson, B., Thomas-Stonell, N., Robertson, B., & Rosenbaum, P. (2019). Validity of a streamlined version of the Focus on the Outcomes of Communication Under Six: Process and outcome. *Child: Care, Health and Development, 45*(4), 600–605. https://doi.org/10.1111/cch.12669

Perry, L. K., Kucker, S. C., Horst, J. S., & Samuelson, L. K. (2022). Late bloomer or language disorder? Differences in toddler vocabulary composition associated with long-term outcomes. *Developmental Science, 26*(4), Article e13342. https://doi.org/10.1111/desc.13342

Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language, and Hearing Research, 51*(2), 394–407. https://doi.org/10.1044/1092-4388(2008/029)

Rescorla, L. A. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17,* 141–150. https://doi.org/10.1002/ddrr.1108

Roberts, M. Y., Curtis, P. R., Sone, B. J., & Hampton, L. H. (2019). Association of parent training with child language development: A systematic review and meta-analysis. *JAMA Pediatrics, 173*(7), 671–680. https://doi.org/10.1001/JAMAPEDIATRICS.2019.1197

Roberts, M. Y., & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology, 20*(3), 180–199. https://doi.org/10.1044/1058-0360(2011/10-0055)

Roberts, M. Y., & Kaiser, A. P. (2015). Early Intervention for toddlers with language delays: A randomized controlled trial. *Pediatrics, 135*(4), 686–693. https://doi.org/10.1542/peds.2014-2134

Rollnick, S., Miller, W. R., & Butler, C. C. (2008). *Motivational interviewing in health care: Helping patients change behaviour.* The Guilford Press.

Singleton, N. (2018). Late talkers: Why the wait-and-see approach is outdated. *Pediatric Clinics of North America, 65*(1), 13–29. https://doi.org/10.1016/J.PCL.2017.08.018

Thomas-Stonell, N., Oddson, B., Robertson, B., & Rosenbaum, P. (2013). Validation of the Focus on the Outcomes of Communication Under Six outcome measure. *Developmental Medicine & Child Neurology, 55*(6), 546–552. https://doi.org/10.1111/dmcn.12123

Thomas-Stonell, N., Robertson, B., Walker, J., Oddson, B., Washington, K., & Rosenbaum, P. (2012). *FOCUS: Focus on the Outcomes of Communication Under Six.* Holland Bloorview Kids Rehabilitation Hospital.

Tosh, R., Arnott, W., & Scarinci, N. (2017). Parent-implemented home therapy programmes for speech and language: A systematic review. *International Journal of Language & Communication Disorders, 52*(3), 253–269. https://doi.org/10.1111/1460-6984.12280

Washington, K. N., Thomas-Stonell, N., McLeod, S., & Warr-Leeper, G. (2015). Outcomes and predictors in preschoolers with speech-language and/or developmental mobility impairments. *Child Language Teaching and Therapy, 31*(2), 141–157. https://doi.org/10.1177/0265659014539689

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50*(6), 1562–1592. https://doi.org/10.1044/1092-4388(2007/106)

Zulkifli, S., Short, K., Kleiman, C., Kidd, J. C., Earley, J., Beckett, S., Descallar, J., & McCabe, P. (2023). Evaluating the dose-response relationship of the number of sessions of "It Takes Two to Talk" in young children with language delay. *International Journal of Speech-Language Pathology, 25*(4), 566–576. https://doi.org/10.1080/17549507.2022.2080270

Authors' Note

Correspondence concerning this article should be addressed to Allison Maier, Kingston, Frontenac, and Lennox & Addington Public Health, 221 Portsmouth Avenue, Kingston, ON, CANADA, K7M 1V5. Email: **allison.maier@kflaph.ca**

Acknowledgments

The authors wish to acknowledge the entire Early Expressions team for their support in implementing the study and reviewing the findings.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

SAC
OAC

Speech-Language &
Audiology Canada

Orthophonie et
Audiologie Canada

Communicating care
La communication à coeur

613.567.9968

1.800.259.8519

1000-1 rue Nicholas St.

Ottawa ON K1N 7B7

www.sac-oac.ca | @SAC_OAC