Statistical properties of auditory behaviour outcome measures for children with hearing loss: a scoping review

Siti Munirah Haris, BSc[1], Nadwah Onwi, MSc[2], Nor Azrita Mohamed Zain, PhD[1], Sarah Rahmat, PhD[1], Basyariatul Fathi Othman, PhD[3]

[1]Department of Audiology and Speech-Language Pathology, Kulliyyah of Allied Health Science, International Islamic University of Malaysia, Kuantan Campus, Pahang, Malaysia, [2]Department of Community Health, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia, [3]Eartistics Hearing and Physiotherapy Centre, Taman Tun Dr Ismail, Kuala Lumpur, Malaysia

ABSTRACT

Introduction: Various evaluation tools have been developed to track the growth of auditory-related behaviours of children with hearing loss during intervention. However, the reliability and validity of currently available outcome measures remain uncertain due to the lack of information on their psychometric properties. A lack of reliable outcome measures may jeopardise intervention quality and affect these children's listening skills progression. This scoping review aims to explore the mechanics of producing or developing an outcome measure either completely new or adapted from the original version that is considered as having robust statistical properties.

Materials and Methods: A scoping review was conducted across four databases (PubMed, ScienceDirect, Scopus and Google Scholar). The included articles were written in English, published between January 2010 and June 2023, and specific to predefined keywords. Two independent reviewers screened and selected the final papers using the PRISMScR checklist. A code framework was created to extract information about the publications and conducted by one reviewer. The results were reported using descriptive statistics and narrative synthesis.

Results: The final analysis were conducted on 22 articles out of 452 articles screened. The review identified seven outcome measures presented in various languages. The outcome measures found were the Auditory Behaviour in Everyday Life (ABEL), Functional Listening Index for Paediatric (FLI-P), Infant-Toddler Meaningful Auditory Integration Scale (IT-MAIS), Integrated Scales of Development (ISD), LittlEARS Auditory Questionnaire (LEAQ), Parent's Evaluation of Aural/Oral Performance in Children (PEACH), Parent's Evaluation of Aural/Oral Performance in Children Diary (PEACH Diary), Teachers' Evaluation of Aural/Oral Performance in Children (TEACH) and Parent's Evaluation of Aural/Oral Performance in Children Plus (PEACH+). A total of 13 studies focused on translating, adapting and validating an outcome measure while the remaining investigations validated either the translated or original version of the outcome measures. All original instruments were developed in English and among Western culture, except for the LEAQ which was designed in the German language and for the German population. The outcome measures identified were translated and adapted into Spanish, Turkish, Persian, Hebrew, Arabic, Malay, Yoruba, Polish, Swedish, Hindi, Portuguese, Kannada and Mandarin.

Conclusion: All studies performed an extensive evaluation of psychometric properties and feasibility studies to produce an excellent quality of auditory-related behaviour outcome measure for clinical use with the intended population. A new outcome measure, FLI-P, was found to be clinically useful for the primary provider of learning to listen and spoken language training for children with hearing impairment in Malaysia, i.e., the speech-language therapists.

KEYWORDS:
Paediatric, young children, infant, hearing impairment, hearing loss, audiological outcome, questionnaire

INTRODUCTION

Conventionally, behavioural or objective audiological assessments through audiometric tests were used to measure the accessibility towards sounds post-hearing intervention. However, these assessments have limitations in explaining how a child with hearing loss uses and integrates the sounds they hear in everyday life.[1,2] Subjective tests, such as questionnaires, diaries and structured interviews, serve as the outcome measures that can address this gap by assessing the auditory-related behaviours in real-world listening environments.[2-4] These measures require the parent(s) and/or teacher(s) observational opinion to quantify a child's auditory or oral performance in everyday listening situation.[2-4] Therefore, tracking a child's skill development and progress using a valid auditory-related behaviour outcome measure is a necessary routine for professionals involved in intervention using spoken language, considering the significant impact of hearing towards language and speech development.[5-6] These professionals include the early interventionists, auditory-verbal therapists, speech language therapists (SLTs), audiologists and/or teachers of the deaf.[5-7]

Standardised and valid auditory-related behaviour outcome measures are important for accurately monitoring the progress and outcomes of children with prelingual hearing loss receiving listening intervention.[7] Numerous auditory

This article was accepted: 15 April 2024
Corresponding Author: Siti Munirah Haris
Email: muniiharis@gmail.com

inventories have been developed for measuring a child's listening skills post-intervention.[3,4,8] Many of them were developed for English-speaking populations, which may not be appropriate for different populations and cultural contexts.[9,10] However, the selection of auditory behaviour measurement tools to assess the intervention outcomes by Malaysian early intervention professionals is influenced primarily by the accessibility, usability, comfort and familiarity of the tools.[1] This scenario results in considerable variability and disparities in reporting outcomes, even among professionals within the same field, particularly among SLTs—the primary providers of early intervention for spoken language in Malaysia.[1] As the evaluation and reporting processes are time-consuming, such evaluation and tracking are not regularly done at specified intervals.[6,7] This imposes significant difficulties for clinicians to track a child's progress and needs for further intervention to improve outcomes. Furthermore, Moodie et al.,[11] emphasised the importance of having an evidence-based age-appropriate outcome measure for effective collaboration between professionals and parents in forming decisions for a child's individualised rehabilitation plan.

Reliable data tracking on functional listening skill progress is fundamental in guiding intervention for better language outcomes while providing support for further decisions and directions in the rehabilitation plan.[1,3,5,6] Therefore, the present scoping review aimed to explore the mechanics of producing or developing an outcome measure either completely new or adapted from the original version that is considered as having robust statistical properties i.e. those with good test-retest reliability, internal consistency, validity and responsivity.[3]

For the purpose of readability of this paper, the authors will use the term auditory-related behaviours to represent behaviours as defined in the Erber's (1984) auditory hierarchy and also more complex spoken language levels that are the consequences of sophisticated auditory functioning.[1]

MATERIALS AND METHODS

Information Sources

A systematic search of four databases, i.e., PubMed, ScienceDirect, Scopus and Google Scholar was performed between 14 June to 21 June 2023 for articles published between published between January 2010 and June 2023. Searched terms and strategies were developed and supported by two researchers in this study. Keywords and related MeSH terms associated with the audiological measure, paediatric, infant, young child, listening outcome, questionnaire, hearing loss and hearing impairment with various combinations were used in the search domains depending on the search settings in selected databases.

Searching Techniques

The scoping review was conducted based on Arksey and O'Malley's five stages of methodological framework.[12] It involved five stages as described below.

Stage 1: Identifying the research questions

1. What are the auditory-related behaviour outcome measures that are available for professionals who provide listening intervention to children with hearing loss all over the world?
2. What are the reliability and validity status as well as the method used in measuring the statistical qualities of the selected auditory-related behaviour outcome measures?

Stage 2: Identifying relevant studies

A systematic electronic search was conducted on four databases, i.e., PubMed, ScienceDirect, Scopus, and Google Scholar between 14 June to 21 June 2023 for articles published between January 2010 and June 2023. Searches on the grey literature with specific keywords of known research studies were conducted to expand the chances to obtain more data. Any unpublished articles or studies regarding auditory-related behaviour outcome measures, as well as non-English published articles were excluded to minimise potential disputes in the reviewed data. The keywords of outcome measures, outcome evaluation, audiological, paediatric, infant, young child, listening outcome, questionnaire, hearing loss and hearing impairment with various combinations were used in the search domains depending on the search settings within the selected databases. The search results were uploaded onto a reference management software and any duplicates were removed. The remaining abstracts were imported into a citation account shared by all researchers.

Stage 3: Study selection

Two reviewers were involved in the screening of studies against the eligibility criteria. All studies included in this review must be published between 2010 to 2023 in the English language. The outcome measures used within this study should meet the following criteria, in which they must: 1) Measure the auditory-related behaviour in real-world listening environments; 2) Provide the psychometric qualities data and 3) Include the skills of children aged between 0 and 6 years. Each study was independently evaluated by both reviewers and an initial screening of titles and abstracts was performed to remove studies that were not within the scope of this review. It was followed by another independent screening and review of the publications' titles, abstracts and full-text copies by both reviewers to eliminate articles that failed to meet the inclusion criteria. Findings from both reviewers were further contrasted during a single discussion conference two weeks after the screening process. A third and fourth reviewer were contacted for further consultation and review of the whole study.

Stage 4: Charting the data

Two evaluators independently reviewed each article and constructed a data extraction form using the Microsoft Excel software (Microsoft, Inc, Redmond, WA, USA). Only articles that satisfied the inclusion criteria were captured in the data extraction form. Following a discussion about whether the charted articles answered the research questions, both evaluators and the third reviewer agreed to finalise the data, which included the author(s), publication year, study location, research objective(s), methodology, subjects/participants, reliability and validity data and conclusion.

Stage 5: Collating, summarising, and reporting results
All articles were gathered, reviewed and reported on the following themes: (1) Study characteristics, (2) Outcome measure description and features, (3) Validity and reliability data and (4) Conclusion of the study. The results section contains a detailed summary of the data acquired from this review. A fifth reviewer was contacted to review the whole report and provide consultation to enhance the readability of the paper.

RESULTS

The literature search generated 545 results across four databases, namely PubMed (239 results), Scopus (65 results), ScienceDirect (180 results) and Google Scholar (61 results). Any duplicates were deleted, resulting in 452 articles related to the research topic that were further screened using the established inclusive criteria. The title and abstract screening of these articles produced 32 articles that were chosen for full-text retrieval; however, three articles were unable to be retrieved. Following the full-text screening, seven articles were judged to be ineligible for inclusion, leaving a total of 22 articles that were included in this scoping review. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 28 was used as the guideline for this scoping study (Figure 1). Different methods, testing procedures and analyses were identified across studies. Therefore, a narrative approach was employed to report the findings of all studies included in the scoping review.

Characteristics of the Studies
The scoping review found that 13 studies were aimed towards translating and adapting the original version of an auditory-related behaviour outcome measures followed by a psychometric properties evaluation of a newly produced measure.[2,4,13-23] The outcome measures identified in these studies were translated and adapted into different languages, including Spanish,[13,16,24] Turkish,14 Persian,[15,22,25] Hebrew,[17] Arabic,[17] Malay,[4,18,23] Yoruba,[19] Polish,[26] Swedish,[27] Hindi,[20] Portuguese,[28] Kannada,[2] and Mandarin.[21,29] All of the original outcome measures were produced in English except for one which was produced in German.[24] The age range of children investigated in these studies was between three months and 18 years of age as the selection of participants was based on the indicated age range of each outcome measure. The characteristics of all studies included in this review are listed in Table I.

The other nine studies targeted towards validating either the original version[6,30,31] or the translated and adapted version of an outcome measure.[14,24-27,29] This review found three out of nine studies had different goals in their validation study which led to different methods and research designs. The first study investigated the feasibility of LittlEars Questionnaire (LEAQ), as a screening tool to identify abnormal hearing development in children, especially in situation where objective measures might not be available.[31] As such, part of the original measure was removed to shorten the measure and to suit the age range of infants involved in the study (i.e., ≤12 months old). The second study aimed to modify their already translated measure, Mandarin version of Infant-Toddler Meaningful Auditory Integration Scale, which was found unsuitable to be used with children who did not receive auditory intervention in their previous study.[29] The investigators used a combination of item response theory (IRT) and classical test theory (CTT) to modify the measure before evaluated its psychometric properties. Another study attempted to modify, validate and compare the new translated version of Persian- Auditory Behavior in Everyday Life Questionnaire with the previously translated version.[25]

Outcome Measures Description and Features
Eight audiological outcome measures were identified from the current review. The Functional Listening Index for Pediatric (FLI-P) by Davis et al., was the newest outcome measure, published in 2022, and employed to measure the auditory-related behaviour of children aged zero to six years old.[6] Six studies were found to use the outcome measure that catered the age range from infancy through childhood, namely the Parents' Evaluation of Aural/Oral Performance of Children Rating Scale (PEACH), Parents' Evaluation of Aural/Oral Performance of Children Diary (PEACH Diary), Parents' Evaluation of Aural/Oral Performance of Children Rating Scale Plus (PEACH Plus)[4,13,14,23,30] and two studies chose the Auditory Behaviour in Everyday Life (ABEL).[25,28] Furthermore, 11 studies used the LittlEARS Auditory Questionnaire (LEAQ),[2,16,17,19-22,24,26,27,31] which was developed for children aged 0 to 24 months. Other outcome measures identified in this review were the Infant-Toddler Meaningful Auditory Integration Scale (IT-MAIS) which was used by one study,[29] the Integrated Scales of Development (ISD) which was used by one study,[18] and the Teachers' Evaluation of Aural/Oral Performance of Children (TEACH) which was also used by one study.[15] IT-MAIS was developed for older infants through childhood years, whilst ISD was designed for children aged 0 to 48 months. Another outcome measure, TEACH, was meant for older children who have attended early intervention centres, preschool or school throughout their childhood years. All outcome measures aimed at measuring the auditory-related behaviour or functional listening skills in real-world listening situations; except for IT-MAIS which only has a few items that cover beyond the sound detection and discrimination level. Additionally, FLI-P, LEAQ and ISD were identified to measure the skills over time and provide the steps for development. However, only FLI-P enlisted the items for measuring advanced auditory-related behaviour skills in older children. The description of outcome measures in this study is provided in Table II following the format presented in Bagatto et al.[3]

Translation and Adaptation Process
All 13 translation studies used the forward-backward translation method with a different number of translators of varying qualifications and backgrounds. At least one translator either from the research team or outside professional was recruited to perform the forward translation. The same scenario was observed in the backward translation process. The harmonisation stage where expert panel reviewed the translated and adapted version was mentioned in all studies with different number and background of panels, except for one study that did not describe this stage in the article.[21] Different methods of validating the content and face of the new measures were also discovered and specifically mentioned only by five studies[14-16,18,22] while other

translation studies did not mention this type of validity stage. Table III outlines the characteristics of the translation and adaptation process for each study in this review.

Validity and Reliability Data

Internal consistency

All thirteen translation studies measured the internal consistency using the Cronbach's alpha value and reported high internal consistency which indicates good reliability of the translated scale measurements. In contrast, only five out of nine validation-only studies – Persian ABEL, Polish LEAQ, Brazilian Portuguese PEACH rating scale, original English PEACH, Mandarin IT-MAIS and Swedish LEAQ – measured internal consistency value using Cronbach's Alpha value[25,26,28-30] while the remaining studies did not mention this value in their report.

Construct validity

Nine translation studies showed good construct validity by reporting high item and total score correlations[2,4,13,14,16,20-23] which indicate a robust positive relationship between an individual item within the test and the overall score of the entire instrument. However, only five out of nine validation-only studies reported the same high construct validity which were calculated using either the regression analysis,[30] factor analysis,[25] item-total correlation,[26,29] and/or difficulty indices.[24,26]

Criterion-related validity

For concurrent validity where the correlation between age and total scores was measured, nine translation studies reported good concurrent validity[2,4,15-17,19-22] while only five validation-only studies reported good concurrent validity.[6,24,27,28,31] These studies found a positive correlation between age and total scores, indicating that older participants tended to score higher on the test. This suggests that the instrument demonstrates good concurrent validity because its scores align well with participants' ages. The elevated sensitivity and specificity were mentioned in three translations[13,14,17] and five validation-only studies6,[25-28] which indicates that the translated instrument is effective in accurately identifying both individuals with and without the hearing impairment. The measurement was performed by measuring the correlations of total scores with different variables (type, degree and laterality of hearing loss, type of device, additional needs, age at device fitting, duration of hearing aids usage before cochlear implantation, chronological age, age at implant activation, age at hearing aid fitting, responses to sounds while using hearing aids before cochlear implantation, and daily usage of hearing aids). High convergent validity was observed in two studies.[27,29] Five studies reported high predictive accuracy in either detecting a child with hearing loss or predicting the future outcome of the auditory-related behaviours of the child.[6,20-22,31]

IV- Test-retest reliability

Out of thirteen translation studies, seven performed test-retest reliability correlation at different intervals ranging from seven days to as long as four weeks and good correlation values were reported by all studies.[2,4,14,15,17,20,23] Five validation-only studies also performed the test-retest reliability with varying number of administration times and intervals with good test-retest reliability.[6,25-27] The number of test administration and duration of test intervals ranged from once in 15 days to 12 in two years (two months interval).[27] High correlations indicate that the measurements taken at different points in time were consistent and stable, suggesting that the tested measure maintained its reliability over time.

V- Index of difficulty

Only two studies measured the difficulty index of their instruments.[21,26] One is the translation study of Mandarin version of LEAQ by Wang et al.,[21] and the other is the validation-only study of Polish version of LEAQ by Obrycka et al.,[26] Both studies reported a good range of difficulty index between 0.31 to 1.00 and 0.52 to 1.00 respectively. The ranges indicated that the items in the instruments were arranged nearly in ascending order of difficulty, with the 'easiest' items representing basic auditory-related behaviour skills and the most 'difficult' ones representing advanced auditory-related behaviour skills.

VI- Normative curves

Eight translation studies plotted the normative curves using the linear regression analysis[4,16,17,19-23] with one study of Hebrew and Arabic version of LEAQ mentioned distinctively about the similarities of their normative curve with the original normative curve using Pearson's correlation.[17] Normative curve was only reported in two validation-only studies in which one study on the original PEACH rating scale by Bagatto et al., compared the produced curve with the previously generated curve[30] and the other study of Swedish LEAQ plotted the predictive growth curve using the linear regression mixed model.[27] Table IV provides an overview of reliability and validity data for all translation studies whilst Table V provides the validity and reliability data of the validation-only studies .

Conclusion of Studies

From the review, 21 out of 22 studies concluded that the outcome measures that were translated, adapted, and/or validated in their studies were reliable and valid in measuring the auditory-related behaviour development of children with specific listening and language environment.[2,4,6,13-17,19-31] They were also useful for clinical practice to monitor and evaluate the effectiveness of auditory-related behaviour skills intervention among children with hearing loss. However, the study of the Malay version of ISD is recommended by the authors to be utilised only as a guide to monitor communication development rather than as an assessment tool.[18] Another study of LEAQ in Swedish mentioned that their studied measure assessed auditory-related behaviours and language skills to a large extent rather than just the audition alone.[27] A study about the suitability of the shortened version of LEAQ as a screening tool was promising as it was easily implementable and served as a good alternative in countries with no objective screening instruments available.[31] Furthermore, a study that modified the existing translated measure found that the combined use of IRT and CTT provided a powerful means to modify psychometrically robust scales.[29] A study which aimed towards validating a newly developed outcome measure, FLI-P, found that the scores derived from their measure can guide

Table I: The characteristics of studies included in the review.

Author	Outcome measure	Country (Language)	Version	Age (Number of subjects)
Bagatto et al.[30 (a)]	PEACH rating scale	Canada (English)	Original	2 to 83 months (n=59 TH)
Bravo-Torres et al.[13 (a)]	PEACH rating scale	Spain (Spanish)	Translation & Adaptation	4 to 18 years (n=297 TH)
Davis et al.[6 (a)]	FLI-P	Australia (English)	Original	0 to 72 months (n=543 HI; 32 TH)
Eroğlu et al.[14 (a)]	PEACH rating scale	Turkiye (Turkish)	Translation & Adaptation	3 to 12 years (n=120 HI)
Fatahi et al.[15 (a)]	TEACH rating scale	Iran (Persian)	Translation & Adaptation	2 to 11 years (n=40 TH; 42 HI)
García et al.[16 (a)]	LEAQ	Spain (Spanish)	Translation & Adaptation	19 to 24 months (n=215 TH)
Geal-Dor et al.[17 (a)]	LEAQ	Israel (Hebrew & Arabic)	Translation & Adaptation	9 to 24 months (n=70 TH (Hebrew) + 97 TH (Arabic); 42 HI)
Hani et al.[18 (a)]	ISD	Malaysia (Malay)	Translation & Adaptation	16 to 30 months (n=12 TH)
Kayode et al.[19 (a,c)]	LEAQ	Nigeria (Yoruba)	Translation & Adaptation	6 to 24 months (n=423 TH)
Obrycka et al.[26 (a)]	LEAQ	Poland (Polish)	Translation & Adaptation	6 to 22 months (n=122 HI)
Oryadi et al.[25 (a)]	ABEL	Iran (Persian)	Translation & Adaptation	1 to 6 years (n=113 HI)
Persson et al.[27 (a)]	LEAQ	Sweden (Swedish)	Translation & Adaptation	16 to 59 months (n=25 TH)
Prakash et al.[20 (a)]	LEAQ	India (Hindi)	Translation & Adaptation	6 to 24 months (n=59 TH; 41 HL)
Quar et al.[4 (a)]	PEACH Diary	Malaysia (Malay)	Translation & Adaptation	3 months to 13 years (n=74 TH)
Quar et al.[23 (a)]	PEACH+ rating scale	Malaysia (Malay)	Translation & Adaptation	4 months to 7 years (n=157 TH)
Levy et al.[28 (a)]	ABEL	Brazil (Portuguese)	Translation & Adaptation	4 to 14 years (n=18 HI)
Schaefer et al.[31 (a)]	LEAQ	German (German)	Original	0 to 60 months (n=47 (6 HI))
Spitzer et al.[24 (a)]	LEAQ	Spain (Spanish)	Translation & Adaptation	5 to 21 months (n=50 TH)
Umashankar et al.[2 (a)]	LEAQ	India (Kannada)	Translation & Adaptation	1 to 24 months (n=67 TH; 20 HI)
Wang et al.[21 (a)]	LEAQ	China (Mandarin)	Translation & Adaptation	4 to 24 months (n=157 TH)
Yang et al.[29 (b)]	IT-MAIS	China (Mandarin)	Translation & Adaptation	0 to 24 months (n=450 TH+HI)
Zarifian et al.[22 (a)]	LEAQ	Iran (Persian)	Translation & Adaptation	Below 24 months (n=240 TH)

(a) Measures auditory-related behaviour skills
(b) Partially measures auditory-related behaviour skills
TH – Typical hearing; HI – Hearing impaired

Table II: The description of outcome measures found in this review.

Outcome measure	Number of items	Response format	Scoring format	Age range	Factors assessed	Developer/Reference Author
ABEL	24	7-point scale	Subscale and overall averages	4 to 14 years	Aural-oral, auditory awareness, social/conversational	Purdy et al.[36]
FLI-P	64	Mostly/Rarely	Total of 'mostly' responses	0 to 6 years	Six phases of auditory behaviours, organised in developmental hierarchy	Davis et al.[4]
IT-MAIS	10 probes	Parental observation and reports via structured interview	Overall score (based on examples given)	Older infancy through childhood	Vocalisation behaviour, alerting to sounds, meaning from sound	Geier[34]
LEAQ	35	Yes/No	Total of 'yes' response	Birth to 24 months	Three categories of auditory behaviours, organised in developmental hierarchy	Kuehn-Inacker et al.[37]
PEACH Diary	13	5-point rating scale	Subscale and overall percentages	Infancy through childhood	Use of devices and loudness discomfort, listening in quiet and noise, phone use, environmental sounds	Ching et al.[35]
PEACH Rating Scale	13	5-point rating scale	Subscale and overall percentages	Infancy through childhood	Use of devices and loudness discomfort, listening in quiet and noise, phone use, environmental sounds	Ching et al.[8]
PEACH+ Rating Scale	12	5-point rating scale	Subscale and overall percentages	Infancy through childhood	Use of hearing devices, listening in quiet situations, listening in noisy situations, ease of demonstrating listening behaviour in different situation	Ching et al.[8]
TEACH	11	5-point rating scale	Subscale and overall percentages	Older children through childhood	Hearing aid use, loudness discomfort, communication in quiet and noise, environmental sounds	Ching et al.[8]

Table III: The translation and adaptation characteristics of each study.

Study-Author	Forward translation	Harmonisation I	Backward translation	Harmonisation II	Content and face validity
PEACH Rating Scale, Bravo et al.[13]	3 native speakers	1 PT, 1 linguist, 1 audiologist. 5 audiologists (different countries; same language)	1 native English Speaker	NA	20 parents
PEACH Rating Scale, Eroğlu et al.[14]	2 audiologists, 1 linguist, 1 PT	2 audiologist, 1 ENT	1 linguist	1 linguist reviewed	40 parents
TEACH Rating Scale, Fatahi et al.[15]	Followed the International Quality of Life Assessment Project Protocols				10 audiologists, 10 teachers
LEAQ, García et al.[16 (a)]	1 PT, 1 psychometrician, 1 SLT	Reviewed with researchers	1 PT, 1 psychometrician, 1 SLT	NA	3 SLT, 30 parents
LEAQ, Geal-Dor et al.[17]	Authors	NA	Authors	1 SLT	NA
ISD, Hani et al.[18]	Authors	3ST, 1 linguist, 1 psychologist, 2 students ST	Authors	10 lecturers SLT	3 parents
LEAQ, Kayode et al.[19]	1 linguist	NA	1 linguist	NA	3 ENT physician
LEAQ, Prakash et al.[20]	4 audiologists	Authors	2 audiologists	Authors	2 SLT
PEACH Diary, Quar et al.[4]	Author	2 audiologists	1 audiologist	NA	6 parents
PEACH+ Rating Scale, Quar et al.[23]	1 English language teacher	Reviewed with researcher (audiologist)	2 translators	1 linguist & 1 audiologist	NA
LEAQ, Umashankar et al.[2]	5 audiologists	Authors	3 audiologists	Authors	3 SLT
LEAQ, Wang et al.[21]	No specific number and flow mentioned				
LEAQ, Zarifian et al.[22 (a)]	1 ST, 1 audiologist, 1 psychologist	Experts review (CVI)	1 SLT	Developer representative	NA

PT – Professional translator, SLT – Speech-language therapist, (a)Cognitive interviewing

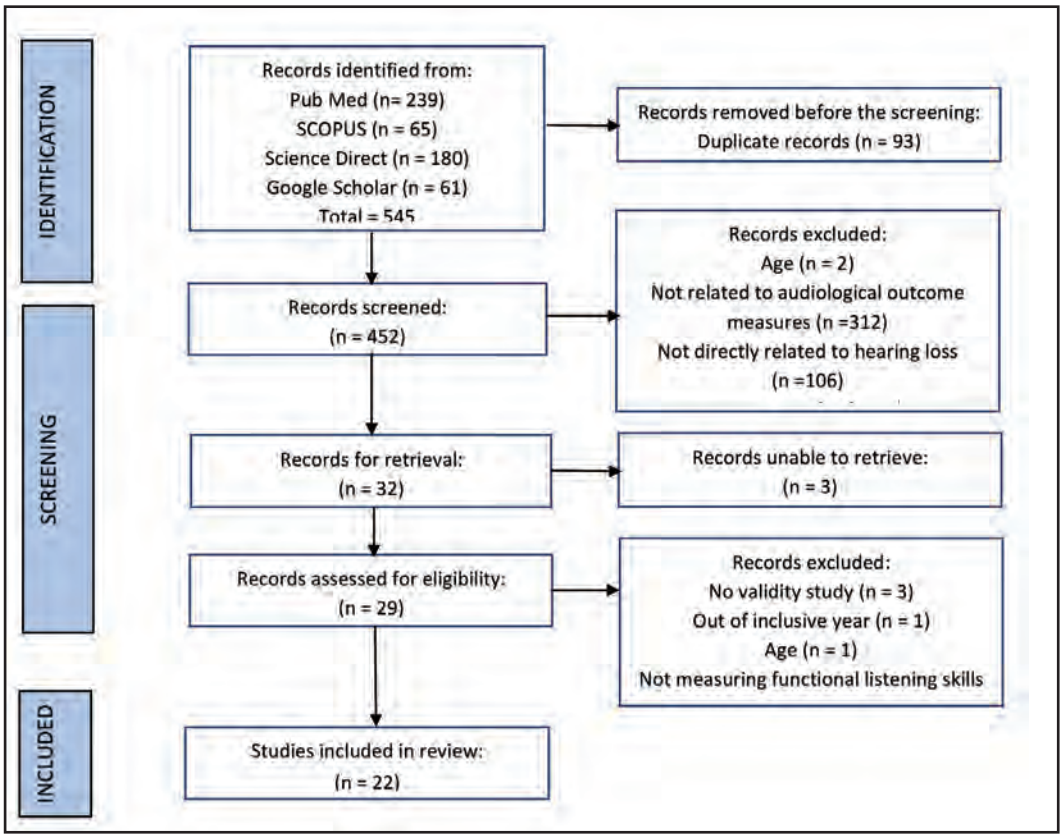

Fig. 1: Flow of the literature search and screening process.

Table IV: The validity and reliability data of translation studies.

Author	Outcome measure	Cronbach's α (Internal consistency)	Test-retest reliability	Validity
Bravo-Torres et al.[13]	PEACH rating scale	0.93	NA	Strong content validity Good construct validity Elevated sensitivity and specificity Good discriminant validity
Eroğlu et al.[14]	PEACH rating scale	0.94	r=0.949 (3 to 4 weeks interval)	Good construct validity High sensitivity and specificity
Fatahi et al.[15]	TEACH rating scale	0.98	r=0.87 to 0.97; (2 weeks interval)	High content validity index Good discriminant validity Good concurrent validity Strong correlation between P-TEACH and P-PEACH
García et al.[16 (a)]	LEAQ	0.93	NA	Good construct validity Good concurrent validity Good Discriminant validity
Geal-Dor et al.[17 (a)]	LEAQ	Hebrew – 0.96 Arabic – 0.95	F (1.09,7.62)=9.468, p=0.015	Good concurrent validity Good convergent validity Good discriminant validity
Hani et al.[18]	ISD	16-18 months – 0.87 19-24 months – 0.63 25-30 months – 0.63	NA	Good content and face validity
Kayode et al.[19 (a)]	LEAQ	0.90	NA	Significant predictive accuracy Good discriminant validity Good concurrent validity
Prakash et al.[20 (a)]	LEAQ	0.96	Z=–1.81, p > 0.05 (7 to 9 days interval)	High predictive validity Good discriminant validity Good concurrent validity
Quar et al.[4 (a)]	PEACH Diary	0.93	χ=- 0.6; range=- 0.7 to - 3.3 (2 weeks interval)	Good construct validity (high item-total correlations)
Quar et al.[23 (a)]	PEACH+ Rating Scale	0.90 (frequency of auditory behaviour), 0.93 (ease of listening behaviour)	t(9)=-1.327 to -0.429, p>0.05) (2 weeks interval)	Good construct validity (high item-total correlations)
Umashankar et al.[2(a)]	LEAQ	0.75	Z=0.94, p > 0.05	Fair concurrent validity Good construct validity
Wang et al.[21 (a)]	LEAQ	0.94	NA	High predictive accuracy Good construct with original validity Good discriminant validity Good concurrent validity Good range of index difficulty
Zarifian et al.[22 (a)]	LEAQ	0.96	NA	Good predictive accuracy Good concurrent validity Good construct validity

[(a)] Has normative data

and support discussion and intervention decisions and also bridge the gap between information from audiological assessments and language measure.[6] Out of all the studies reviewed, nine have produced normative curves in which the professionals may plot on to track the child's development of the auditory-related behaviours.[4,16,17,19-22,27,30]

DISCUSSION

The Available Auditory-Related Behaviour Outcome Measures

This scoping review found eight different instruments in various languages that are useful for measuring the auditory-related behaviour development, and subsequently monitoring the intervention outcome in children with hearing loss. Although these outcome measures differ in their features and clinical indications, all of them demonstrated good reliability and validity values, indicating their ability to measure what it is supposed to measure with consistent findings and results. In general, the knowledge about clinical features of each outcome measure is important in guiding the clinicians to choosing and determining the most feasible and viable instrument for their clinical use.[3] The PEACH Diary, PEACH Rating Scale, PEACH+ Rating Scale, IT-MAIS, LEAQ and FLI-P were identified to be appropriate to measure the auditory-related behaviours from infancy, with LEAQ narrowing its focus of development only up to 24 months old.[31] The ABEL and TEACH on the other hand, were more suitable for pre-school children, preferably from 4 years old. The PEACH Rating Scale, PEACH+ Rating Scale, LEAQ and FLI-P showed good clinical feasibility and responsivity due to their administration via interview-observation or past-self-recollection, rated by either the clinicians or parents. In contrast, for both PEACH Diary and IT-MAIS, consideration for their practicality and ease of use has to be made because

Table V: The validity and reliability data of validation-only studies.

Author	Outcome measure	Cronbach's α (Internal consistency)	Test-retest reliability	Validity
Bagatto et al.[30 (a)]	PEACH rating scale	0.78	NA	Good construct validity Good concurrent validity High sensitivity and specificity
Davis et al.[4]	FLI-P	NA	Steep inclines in listening trajectories over time (3 to 4 months interval)	Good construct validity Good concurrent validity Good predictive validity Good discriminant validity High sensitivity and specificity
Obrycka et al.[26]	LEAQ	0.83	Significance difference in auditory development (Test intervals – 1,3,6,9,12 months old)	Good construct validity Good range of difficulty index Good concurrent validity
Oryadi et al.[25]	ABEL	0.96	df=5, F=35.67, p value < 0.001	Good construct validity
Persson et al.[27 (a)]	LEAQ	NA	Every 2 months for 1 year F (3.894, 93.467) =368.304, p<0.001 Every 2 to 4 months for 2 years	Good convergent validity - LEAQ and McArthur-Bates CDI. Weak to no correlations with PEACH LEAQ. Measure language skill rather than audition.
Levy et al.[28]	ABEL	>0.7	p>0.05 (no significant difference) (15 days interval)	Good concurrent validity Poor to negative sensitivity
Schaefer et al.[31]	LEAQ	NA	NA	Good predictive accuracy for detecting hearing loss Low predictive accuracy for detecting speech delays and language development
Spitzer et al.[24 (a)]	LEAQ	0.95	NA	Good construct validity Good criterion validity Good discriminant validity
Yang et al.[29]	IT-MAIS	0.92	NA	High convergent validity Good construct validity Good concurrent validity

[(a)] Has normative data

of their open-interview style administration format as described in the review by Bagatto et al.[3] Another useful features for a clinician when tracking a child's skill developmentally is the normative curve, that is represented as the trajectory graph of scores. These graphs which are included in PEACH, PEACH+ Rating Scale and FLI-P, provide an over-time tracking framework to the intervention team for informed decision making and determining intervention direction. Further scrutiny revealed that FLI-P is the only outcome measure that enlists a wide range of auditory-related behaviour skills in a real-world listening environment. FLI-P sequences these skills hierarchically, starting from early sound awareness phase up to advance open-set phase, making it extensively different from the other measures in this review.[6] This plus point feature was found to give greater impact clinically in guiding SLTs to set the intervention aims, and discuss the intervention outcome with the parents and other professionals on the intervention team. Another almost similar measure to FLI-P is the ISD. The ISD adopts the milestone checklist-like presentation for five different developmental areas simultaneously,[18] rendering a much simpler and generalized auditory-related behaviours checklist compared to the FLI-P. This feature reduces ISD practicality in monitoring a child's progress clinically as well as in setting therapy focus.

The Validity and Reliability Status and Methods in Measuring the Statistical Properties

The selection of outcome measures for translation and adaptation is determined by the clinical indication of the population in addition to the validity and feasibility evidence provided by the developer.[33] As recommended by Hall et al,[32] investigators should choose a measure that requires minimal changes with relevant and equivalent concept of interest across sources and target countries where it will be used for translation and adaptation studies. Despite the variations in the types of validity and reliability measurements reported by each study, the primary focus remains on validating the newly developed instrument to ensure its clinical feasibility for the targeted population. Majority of the studies reviewed consistently reported the internal consistency values of their measures. Internal consistency indicates the strong reliability of the translated scale measurements.[3] All translation study reported high Cronbach's Alpha value for their internal consistency measurement whilst only a few of the validation-only studies did. This discrepancy in reporting highlights the importance of consistently evaluating and reporting internal consistency across studies to ensure transparency and reliability in the assessment of measurement instruments. Construct validity assesses the degree to which a measurement tool accurately measures the underlying construct or concept it is intended to measure.[33] Typically, it

is evaluated through various methods such as factor analysis, convergent validity and discriminant validity. Almost all studies in this review reported construct validity by measuring the item and total score correlations except for a few validation-only studies that did not perform this type of validity measurement. This suggests variations in the validity of the instruments across different contexts or populations which shall be taken into consideration when used clinically. Criterion-related validity was another type of validity observed in the reviewed studies. One of them was the concurrent validity which was assessed through the correlation between age and total scores of an instrument. Additionally, elevated sensitivity and specificity were also observed in several studies, which indicated the effectiveness of the translated instruments in accurately identifying individuals with and without hearing impairments. These findings underscore the importance of assessing multiple aspects of criterion-related validity to ensure the accuracy and effectiveness of the measurement instruments. Another statistical measure found in the reviewed studies was the test-retest reliability. Majority of studies demonstrated good test-retest reliability, with consistent and stable correlations observed over varying intervals. This indicates that the measurements taken at different points in time were reliable and consistent, suggesting that the instruments maintain their reliability over time. The majority of translation studies plotted normative curves using linear regression analysis, providing valuable reference points for interpreting scores in clinical settings. However, normative curves were less frequently reported in validation-only studies. This suggests that while translation studies focus on establishing normative data for the translated instruments, validation-only studies may prioritize other aspects of reliability and validity of their assessments.

LIMITATIONS

This scoping review was conducted based on the PRISMA statement with a comprehensive literature search strategy. However, several exclusion criteria included during the searches may have inadvertently led to the exclusion of some prominent and relevant research studies. These include the limitation of publication years, the exclusion of non-published literature, publication in a non-English language, and those that could not be retrieved in full article.

CONCLUSION

This current review discovered that all studies performed an extensive evaluation of psychometric properties and feasibility studies to produce an excellent quality of auditory-related behaviour outcome measure for clinical use with the intended population. In summary, the findings from both translation and validation-only studies provide strong support for the reliability and validity of the instruments for assessing listening intervention outcomes. Although some measures were modified to suit the target population, the studies were able to prove consistent reliability and validity outcomes which were comparable with the original measures. However, variations in reporting and the assessment of certain validity aspects across studies highlight the need for standardized methods and transparent reporting practices in future research. This is to ensure the robust and reliable outcome measure instruments in both clinical and research settings. In addition, this review also found a high potential outcome measure, FLI-P, which has strong constructs and practical usability especially for the SLTs who provide learning to listen and spoken language training for children with hearing impairment in Malaysia. It is recommended that future studies attempt to translate and validate some of the reviewed outcome measures into Malaysian main languages (Malay, Mandarin and Tamil) for research and clinical usage.

CONFLICT OF INTEREST AND FUNDING

The authors declare no conflict of interest. This research was funded by the Ministry of Higher Education of Malaysia via the Federal Training Award.

REFERENCES

1. Davis A. Improving communication outcomes for children with hearing loss in their early years: Tracking progress and guiding intervention. PhD Thesis. Macquarie University 2020.
2. Umashankar A, Thejas BT, Prabhu P. Translation and validation of the LittlEars auditory questionnaire in Kannada. Int J Pediatr Otorhinolaryngol 2021; 142: 110598.
3. Bagatto MP, Moodie ST, Seewald RC, Scollie SD. A Critical Review of Audiological Outcome Measures for Infants and Children. Trends Amplif 2011; 15(1): 23-33.
4. Quar TK, Ching TYC, Mukari SZ, Newall P. Parents' Evaluation of Aural/Oral Performance of Children (PEACH) scale in the Malay language: Data for normal-hearing children. Int J Audiol 2012; 51(4): 326-33.
5. Coninx F, Weichbold V, Tsiakpini L, Autrique E, Bescond G, Tamas L, et al. Validation of the LittlEARS® Auditory Questionnaire in children with normal hearing. Int J Pediatr Otorhinolaryngol 2009; 73(12): 1761-8.
6. Davis A, Harrison E, Cowan R. The Feasibility of the Functional Listening Index—Paediatric (FLI-P®) for Young Children with Hearing Loss. J Clin Med 2022; 11(10): 2764.
7. Henderson L, Wallis A, de Kruijf M, Fullwood C, Moualed D, Bruce IA, et al. Validation of the Manchester spoken language development scale (MSLDS). Cochlear Implants Int 2020; 21(5): 239-45.
8. Ching TYC, Hill M. The Parents' Evaluation of Aural/Oral Performance of Children (PEACH) scale: normative data. J Am Acad Audiol 2007; 18(3): 220-35. Retrieved from the LOCHI study website https://www.outcomes.nal.gov.au/peach
9. Levinger M, Ronen T. Is it really clear?: adapting research tools for the needs of the deaf population. J Soc Work Pract 2008; 8(4): 399-430.
10. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000; 25(24): 3186-91.
11. Moodie ST, Bagatto MP, Miller LT, Kothari A, Seewald R, Scollie SD. An integrated knowledge translation experience: use of the Network of Pediatric Audiologists of Canada to facilitate the development of the University of Western Ontario Pediatric Audiological Monitoring Protocol (UWO PedAMP v1.0). Trends Amplif 2011; 15(1): 34-56.
12. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005; 8(1): 19-32.
13. Bravo-Torres S, Fuentes-López E, Guerrero-Escudero B, Morales-Campos R. Adaptation and validation of the Spanish version of the Parents' Evaluation of Aural/Oral Performance of Children (PEACH) rating scale. Int J Audiol 2020; 59(8): 590-7.

14. Eroğlu K, Kamışli GİŞ, Altınyay Ş, Gökdoğan Ç, Bayramoğlu İ, Kemaloğlu YK. Validation of the Turkish Version of the Parents' Evaluation of the Aural/Oral Performance of Children (PEACH) rating scale. ENT Updates 2021; 11(3): 165-73.
15. Fatahi F, Hajisadeghian N, Hajiabolhassan F, Abdollahi FZ, Jalaie S. Development of Persian version of teachers' evaluation of aural/oral performance of children scale . Aud Vestib Res 2020; 29(2): 64-75.
16. García Negro AS, Padilla García JL, Sainz Quevedo M. Production and evaluation of a Spanish version of the LittlEARS® Auditory Questionnaire for the assessment of auditory development in children. Int J Pediatr Otorhinolaryngol 2016; 83: 99-103.
17. Geal-Dor M, Jbarah R, Meilijson S, Adelman C, Levi H. The Hebrew and the Arabic version of the LittlEARS® Auditory Questionnaire for the assessment of auditory development: results in normal hearing children and children with cochlear implants. Int J Pediatr Otorhinolaryngol 2011; 75(10): 1327-32.
18. Hani F, Antoinette H, Vandort S, Rahayu, Kamal M. [Adaptation of the integrated scales of development for normal children 16-30 months]. Jurnal Sains Kesihatan Malaysia 2010; 8(2): 9-14.
19. Kayode O, Adeyemo AA. The Yoruba version of LittlEARS Auditory Questionnaire: Evaluation of auditory development in children with normal hearing. J Otol 2018; 13(3): 92-6.
20. Prakash P, Lakshmi S, Sreedhar A, Mathur AV, Konadath S. Validation of LittleEARS questionnaire in Hindi language. J Otol 2023; 18(2): 71-8.
21. Wang L, Sun X, Liang W, Chen J, Zheng W. Validation of the Mandarin version of the LittlEARS® Auditory Questionnaire. Int J Pediatr Otorhinolaryngol 2013; 77(8): 1350-4.
22. Zarifian T, Movallali G, Fotuhi M, Harouni GG. Validation of the Persian version of the LittlEARS ® auditory questionnaire for assessment of auditory development in children with normal hearing. Int J Pediatr Otorhinolaryngol 2019; 123: 79-83.
23. Quar TK, Ooi GHT, Umat C, Mazlan R, Chong FY, Ching TYC. A validation and normative study of the Parents' Evaluation of Aural/Oral Performance of Children plus (PEACH+) rating scale in Malaysia. Int J Audiol 2023; 63(4): 286-91.
24. Spitzer JB, Zavala JS. Development of Spanish Version of the LittlEARS Parental Questionnaire for use in the United States and Latin America. Audiol Res 2011; 1(2): e31.
25. Oryadi-Zanjani MM, Vahab M, Purdy SC. The Persian Version of the Auditory Behavior in Everyday Life Questionnaire. Int J Sch Health 2018; 5(1): 1-5.
26. Obrycka A, Lorens A, Padilla García JL, Piotrowska A, Skarzynski H. Validation of the LittlEARS Auditory Questionnaire in cochlear implanted infants and toddlers. Int J Pediatr Otorhinolaryngol 2017; 93: 107-16.
27. Persson A, Miniscalco C, Lohmander A, Flynn T. Validation of the Swedish version of the LittlEARS® Auditory Questionnaire in children with normal hearing–a longitudinal study. Int J Audiol 2019; 58(10): 635-42.
28. Levy CCA da C, Rodrigues-Sato LCCB. Questionnaire validation - PEACH on Brazilian Portuguese. Codas 2016; 28(3): 205-11.
29. Yang F, Zhao F, Zheng Y, Li G. Modification and verification of the Infant-Toddler Meaningful Auditory Integration Scale: a psychometric analysis combining item response theory with classical test theory. Health Qual Life Outcomes 2020; 18(1): 367.
30. Bagatto MP, Scollie SD. Validation of the Parents' Evaluation of Aural/Oral Performance of Children (PEACH) Rating Scale. J Am Acad Audiol 2013; 24(2): 121-5.
31. Schaefer K, Coninx F, Fischbach T. LittlEARS auditory questionnaire as an infant hearing screening in Germany after the newborn hearing screening. Int J Audiol2019; 58(8): 468-75.
32. Hall DA, Zaragoza Domingo S, Hamdache LZ, Manchaiah V, Thammaiah S, Evans C, et al. A good practice guide for translating and adapting hearing-related questionnaires for different languages and cultures. Int J Audiol 2018; 57(3): 161-75.
33. Prinsen CAC, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a 'Core Outcome Set' - a practical guideline. Trials 2016; 17(1): 449.
34. Geier AK. Development of the Hearing Aid Benefit Scale for Infants and Toddlers. Unpublished doctoral dissertation, Central Michigan University, Mount Pleasant 1998.
35. Ching TY, Hill M. The Parents' Evaluation of Aural/Oral Performance of Children (PEACH) Rating Scale. Chatswood, New South Wales, Australia: Australian Hearing 2005. Retrieved from the LOCHI study website https://www.outcomes.nal.gov.au/peach
36. Purdy SC, Farrington DR, Moran CA, Chard LL, Hodgson SA. A parental questionnaire to evaluate children's auditory behavior in everyday life (ABEL). American Journal of Audiology 2002. 11, 72-82.
37. Kuehn-Inacker H, Weichbold H, Tsiakpini L, Coninx F, D'Haese P. LittlEARS Auditory Questionnaire Manual – Parent Questionnaire to Assess Auditory Behavior in Young Children. MED-EL; Innsbruck, Austria: 2003.