BMJ Open

Study protocol for the validation of a new patient-reported outcome measure (PROM) of listening effort in cochlear implantation: The Listening Effort Questionnaire-Cochlear Implant (LEQ-CI)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-028881
Article Type:	Protocol
Date Submitted by the Author:	31-Dec-2018
Complete List of Authors:	Hughes, Sarah; Swansea University, School of Medicine; Abertawe Bro Morgannwg University Health Board, South Wales Cochlear Implant Programme Rapport, Frances ; Macquarie University, Australian Institute of Health Innovation Watkins, Alan; Swansea University, College of Medicine Boisvert, Isabelle; Macquarie University, Department of Linguistics (Audiology Section); Australian Hearing Hub, The HEARing CRC McMahon, Catherine; The HEARing CRC; Macquarie University, Department of Linguistics (Audiology Section) Hutchings, Hayley; Swansea University, Medicine
Keywords:	hearing loss, cochlear implant, Audiology < OTOLARYNGOLOGY, patient-reported outcome measure, validation, listening effort

Study protocol for the validation of a new patient-reported outcome measure (PROM) of listening effort in cochlear implantation: The Listening Effort Questionnaire-Cochlear Implant (LEQ-CI)

Sarah E. Hughes[1,2]

Frances Rapport[3,6]

Alan Watkins[2]

Isabelle Boisvert[3,4,5]

Catherine M. McMahon[3,4,5]

Hayley A. Hutchings[2]

[1]South Wales Cochlear Implant Programme, ABMU Health Board, Bridgend, Wales, United Kingdom

[2]Patient and Population Health and Informatics, Swansea University Medical School, Swansea, Wales, United Kingdom

[3]Centre for Implementation of Hearing Research, Macquarie University, Sydney, Australia

[4]Department of Linguistics (Audiology Section), Macquarie University, Sydney, Australia

[5]The HEARing Cooperative Research Centre, Australia

[6]Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Corresponding Author:

Sarah E. Hughes

South Wales Cochlear Implant Programme

Princess of Wales Hospital

Coity Road, Bridgend

Wales, United Kingdom CF31 5EU

E-mail: sarah.hughes@wales.nhs.uk

Phone: +44 1656 752192

Fax: +44 1656 752192

Word count (excluding title page, abstract, references): 3239

Key words: hearing loss, cochlear implant, listening effort, patient-reported outcome measure, validation, audiology

ABSTRACT

Introduction: Hearing loss (HL) affects 11 million people in the UK. Associated with HL is a requirement for high levels of effort when listening. Listening effort (LE) may be defined as the attentional and cognitive resources needed to listen. A number of hearing loss-specific patient-reported outcome measures (PROMs) are used routinely in the audiology/cochlear implant (CI) clinic; however, none adequately address the requirement for LE and the impact sustained effortful listening has on individuals' quality of life. This UK-based study aims to undertake an initial psychometric validation of a newly-developed, disease-specific PROM of LE, the Listening Effort Questionnaire-Cochlear Implant (LEQ-CI), in the UK population of adult cochlear implant candidates and recipients. The study will also establish whether further refinements are required to improve its measurement properties.

Methods and analysis: This UK-based study is a multi-phase validation study that has been designed in accordance with the internationally recognised COSMIN standards. In Phase 1, adult CI candidates and recipients (n = 250) will self-complete a paper-and-pencil version of the draft LEQ-CI. Participants' responses to the LEQ-CI's items will be assessed to establish unidimensionality and Rasch Analysis will evaluate item and scale functioning. Classical Test Theory (CTT) will assess acceptability/data completeness, scaling assumptions, targeting, and internal consistency reliability. Phase 1 results will inform adjustments to the items, scale(s), and response options to produce a refined version of the LEQ-CI. In Phase 2, a new sample (n = 100) will self- complete the refined LEQ-CI and four comparator PROMs to assess construct validity.

Ethics and dissemination: This study has been approved by the Abertawe Bro Morgannwg University (ABMU) Health Board/Swansea University Joint Study Review Committee (JSRC) and the Newcastle and North Tyneside 2 Research Ethics Committee (REC), Ref: 18/NE/0320.

Findings will be disseminated in high-quality peer-reviewed journals, conference presentations, and SEH's doctoral dissertation. (299 words)

ARTICLE SUMMARY

Strengths and limitations

- The LEQ-CI is the first PROM developed specifically to assess perceived listening effort in cochlear implant candidates and recipients.
- The proposed study conforms to international consensus standards on best practice of studies of instrument development and validation, the COsensus-based Standards for the selection of health-status Measurement INstruments (COSMIN).
- The use of CTT and Rasch Analysis will enable a robust initial assessment of the LEQ-CI's measurement characteristics at both item and scale level.
- The conceptual framework underpinning the LEQ-CI is based on an explanatory model developed from current theoretical frameworks and the patient perspective. Assessment of the LEQ-CI's measurement properties will provide early evidence of the validity of the proposed model.
- Instrument validation is an iterative process to build a body of evidence relating to the quality of an instrument's measurement properties. Further studies that assess the measurement characteristics of LEQ-CI will be required.

INTRODUCTION

Patient-reported outcome measures (PROMs) are self-report instruments that assess an individual's perception of their disease severity and/or quality of life or well-being.[1–3] As healthcare moves towards more patient-centred models of healthcare delivery, PROMs are being used increasingly in routine clinical practice [4] and are already well-established in the field of audiology.[5] PROMs enable clinicians to gain insight into the patient's perspective of their condition and the treatment they receive. Importantly, PROMs provide insight into those aspects of a disease or condition that are not observable, but rather, are knowable only to the patient themselves.

In the context of hearing loss and audiological clinical practice, many current routine assessments are capable of providing insight into audibility of the acoustic signal but are unable to supply information relating to the underlying processes and mechanisms that inform the measured performance. Listening effort, which may be defined as the attentional and cognitive resources needed to understand an auditory process,[6] is one such factor known to impact on the everyday listening activities of adults with hearing loss with negative implications for physical, mental, and social well-being.[7–10] There is growing interest in perceived listening effort from the research and clinical communities; however, current clinical tools are, as yet, unable to reliably evaluate listening effort and its impact on the listening activities of everyday life.[11,12] As the number of adults with significant HL increases and recognising the significant impact hearing loss will have on these individuals' quality of life and well-being, well-validated measures to assess the underlying factors considered to contribute to an individual's experience of hearing loss will be required.

In the published literature, listening effort has been measured using physiological measures such as pupilometry,[13,14] functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) [15,16] as neurophysiological correlates of listening effort.

Behavioural performance measures (e.g., the dual-task paradigm) as measures of listening effort are also well-reported in the literature.[17] However, the complex nature of the listening effort construct means these tools may not be capturing all aspects of the construct and may not be appropriate for assessing those properties of listening effort that are important to individuals when listening in the real world. A PROM has the potential to offer a clinically viable and alternative view on listening effort from the patient perspective. A systematic review of PROMs considered to measure perceived listening effort established that a number of self-report questionnaires have been developed that include individual items considered to assess listening effort but, notably, found no validated self-report instruments developed specifically to measure perceived listening effort in a manner consistent with current theoretical frameworks.[12,18]

AIM

Health instrument validation is an iterative process whereby evidence of a PROM's psychometric qualities is established in multiple studies over time.[19] The aim of this study is to conduct an initial psychometric validation of a new PROM that has been developed specifically to assess perceived listening effort in the population of adults with severe-profound HL who are CI candidates or recipients. The study builds on previous work undertaken by the authors to establish the content validity of this new instrument, the Listening Effort Questionnaire – Cochlear Implant (LEQ-CI).[20] The current study represents a further step towards the provision of a robust measure of perceived listening effort for use in research and clinical practice.

OBJECTIVES

- To refine the items, response categories, and scale structure of the new LEQ-CI using Rasch Measurement Theory in an English-speaking sample of adult cochlear implant candidates and recipients in the UK.

- To undertake an initial assessment of the LEQ-CI's psychometric properties, applying CTT to assess acceptability, scaling assumptions, targeting, and reliability.
- To assess the construct validity of the LEQ-CI in the population of adults with severe-profound, post-lingual HL who are CI candidates or recipients.

METHODS AND ANALYSIS

Study Setting and Patient Involvement

This study is a UK-based multi-phase study to validate the measurement properties of a new PROM, the LEQ-CI. The planned study will take place over a 12-month period and has been co-produced by the study team with input from two lay members, both CI recipients, from the study's Research Management Group. Lay members reviewed and provided feedback on study design, participant documents and iterations of the LEQ-CI.

Development of the LEQ-CI

The LEQ-CI is a disease-specific PROM measuring perceived or self-reported listening effort in adult cochlear implant candidates and recipients. Item content was developed from the results of a qualitative mixed-methods study [20] with further items harvested from extant PROMs considered to measure listening effort or associated constructs.[18,21] Preliminary testing of the LEQ-CI included assessment of item quality using the on-line Survey Quality Predictor system (Version 2.1, http://sqp.upf.edu/), expert review from a panel of academics, researchers, and clinicians (n = 7), and a series of cognitive interviews to elicit feedback on the relevance, clarity and acceptability of the LEQ-CI from a purposive sample of cochlear implant candidates and recipients (n = 12). After preliminary testing, the LEQ-CI comprises 27 items across four domains. Seven-point or nine-point Likert scales with absolute anchors and labelled

categories ensure a broad range of response options. Item responses are summed to produce a simple total score.

Sample size

The study sample will be representative of the population of adults with acquired, post-lingual severe-to-profound sensorineural hearing loss (SNHL) referred for cochlear implantation in the UK. A total study sample of 350 participants will be recruited. In Phase 1, a cohort of 250 participants will be recruited from multiple National Health Service (NHS) cochlear implant centres. There are no general criteria for determination of sample sizes in studies of PROM validation and sample sizes are, in part, dependent on the psychometric characteristics being assessed.[22,23] Mokkink et al. recommend greater than 200 respondents when undertaking RA and seven times the number of items for purposes of undertaking assessment of unidimensionality.[22] Linacre recommends a sample size of 250 respondents for definitive item calibration using RA.[24] The LEQ-CI is comprised of 27 items that have been selected to minimise respondent burden whilst allowing for adequate sampling of relevant constructs associated with listening effort. Therefore, a minimum sample of 250 participants is considered sufficient for undertaking both assessment of unidimensionality and RA of the LEQ-CI. In Phase 2, a new cohort of 100 participants fulfilling the same eligibility criteria as Phase 1 will be recruited. Hobart et al. recommend greater than 80 participants for assessment of construct validity.[25]

Recruitment and data collection

The participant eligibility criteria are the same for both phases of the study and are presented in Table 1.

Table 1. Study eligibility criteria for recruitment of participants

Inclusion Criteria	Exclusion Criteria
• Adults (persons ≥ 18 years of age). • Post-lingual severe-profound SNHL • A candidate for cochlear implantation according to UK criteria specified by the National Institute of Health and Care Excellence (NICE, 2009) or cochlear implant recipient.[26] • Proficient readers/writers of English. • Capacity to give informed consent. • No additional medical conditions precluding the participant's ability to self-complete the questionnaires	• Children (persons < 18 years of age) • Normal hearing (NH) or a SNHL that does not meet the NICE candidacy criteria for cochlear implantation (e.g., mild-moderate SNHL, or high speech recognition performance with hearing aids).[26] • Pre-lingual severe-to-profound SNHL (i.e., when the onset of the hearing loss can reasonably be estimated to have occurred before age 3, in both ears) and the individual's primary mode of communication is manual (e.g., British Sign Language). • Does not have capacity to give informed consent. • Unable to read/write in English.

In Phases 1 and 2, participants meeting the study inclusion criteria will be sent an invitation letter, an information sheet describing the study in detail, the LEQ-CI, a demographic questionnaire, and comparator questionnaires (Phase 2 participants only). A reply-paid envelope for the return of the completed questionnaires will be provided. Informed consent is presumed if the questionnaires are completed and returned to the study team. To maintain

participant anonymity, eligibility screening and the study documents will be mailed to prospective participants by a member of the clinical team at each participating cochlear implant centre. To maintain participant anonymity, each questionnaire pack will be coded with a unique identifier. No personal identifiable information will be retained by the study team.

Statistical Analysis

Study data will be managed using the online clinical data management programme, REDCap (Version 7.2.1, Vanderbilt University), licensed to the Swansea Trials Unit, Swansea University. RA will be used to assess item and scale structure of the LEQ-CI using Winsteps (Version 4.1.0) software. Psychometric analyses applying CTT will be conducted using the Statistical Package for Social Sciences (SPSS), Version 22.0 licensed to Swansea University.

Phase 1: Item & Scale Refinement using Rasch Measurement Theory

As a form of modern measurement theory, RA is being applied increasingly to the development of PROMs as a complement to CTT, the traditional method of psychometric evaluation. Used to improve the precision of assessment instruments.[27] RA allows instrument designers to empirically assess the behaviour of both items and response categories. Central to the Rasch model is a method for ordering persons (i.e., patients) according to the amount of the latent target construct (i.e., listening effort) they possess and for ordering items that measure the target construct according to their difficulty.[28] This method allows non-linear (ordinal) raw data to be converted to a linear scale (interval), which can then be evaluated through the use of parametric statistical tests.[27] The Rasch model considers how well the observed data fits the measurement model, unlike CTT which considers how well the model describes the data.[29] Because the Rasch model is based on theory and is independent of any data set, any discrepancies between the scale data and the Rasch model requirements are

indicative of anomalies in the scale as a measurement instrument. These discrepancies provide diagnostic information that serves as a basis for understanding and empirical improvement of the instrument at both item and scale-level.[30]

Assessing unidimensionality

The Rasch measurement model assumes unidimensionality, defined as the measurement of a single latent construct.[28,31] Therefore, prior to undertaking RA, factor analysis will be undertaken to assess the underlying structure of the LEQ-CI and establish the unidimensionality of its (sub)scales.[32]

Assessing item fit

In RA, item fit refers to the degree of mismatch between the pattern of actual observed responses and the Rasch modelled expectations. Specifically, it refers to the pattern for each item across persons investigated by examining item infit and outfit statistics.[28] Mean square standardized residuals (MNSQ) will be used to assess fit with MNSQ residuals within the 0.5–1.5 range considered acceptable for productive measurement. Mean square values less than 0.5 indicate overfit (i.e., the items are too predictable relative to the Rasch model), while mean square values greater than 1.5 are indicative of too much noise (randomness) relative to the Rasch model.[33]

Assessing differential item functioning (DIF)

DIF is an indication of the loss of invariance across subsamples of respondents. The presence of DIF will be an indicator of potential problems with an item since item and person measures on a unidimensional instrument should remain invariant (i.e., within error) across all appropriate measurement conditions.[28] DIF will be examined for key demographic variables

such as age and sex. The standard threshold of > 1 logit will be used as an indicator of DIF.[31] The methodology proposed by Zumbo using logistic regression DIF tests for significance (i.e., Chi-squared two degrees of freedom test) and magnitude of DIF by computing the R-squared effect for both uniform and non-uniform DIF will be applied.[34] If items are found to have DIF they will either be considered candidates for removal or examined for adjustment of DIF and re-evaluated, thus reflecting the iterative nature of instrument validation.[28,35]

Assessing response scale ordering

The response options of an instrument (i.e., number of categories and their definitions) are critical to its reliability and validity.[36] The Rasch model will enable us to show empirically how respondents use the LEQ-CI's rating scale informing future iterations of the LEQ-CI to ensure it yields high quality data.[28] Response category ordering will be assessed using Rasch probability curves and there will be an examination of the data for category disordering and threshold disordering.[37] These investigations will show whether the response options selected for the LEQ-CI are sufficient or should be collapsed to provide better coverage of the latent trait.

Assessing the targeting of persons and items

Targeting explores whether the instrument has a distribution of items that matches the range of the respondents' latent trait. This will be done by examining the item-person threshold distribution map, which illustrates a relative position of "item difficulty" to "person ability".[31] The means and standard deviations of items and persons should match closely.[28]

Assessing reliability

The reliability of the LEQ-CI will be examined by observing the person separation index (PSI). The PSI is an estimate of the spread or separation of persons on the measured variable [28] and is considered to be a measure of internal consistency reliability.[38] It is a measure of the scale's ability to separate the study sample. A PSI > 0.7 will be considered an adequate measure of reliability.[39]

Phase 1: Psychometric evaluation using Classical Test Theory

Initial assessment of the LEQ-CI's psychometric properties will be undertaken applying principles of CTT that comply with the current "gold standard" proposed by the COSMIN group.[22] As a theory of measurement, CTT seeks to evaluate the reliability and validity of a scale and has been the dominant approach used in the development and validation of outcome measures. CTT is based on the assumption that every observed score is a function of an individual's true score and random error.[40] The assumptions underpinning CTT differ from those underpinning the Rasch model. Notably, CTT focuses only on the total test score (i.e., the summary of item scores) to establish reliability and validity. Item-level measurement properties are not considered; therefore, limiting the use of CTT as a method to assess the performance of individual items.[41] For this study, item function will be assessed first using RA followed by complementary psychometric analyses using CTT. Specifically, CTT will be used to evaluate the LEQ-CI for its acceptability, targeting, scaling assumptions, and internal consistency reliability.

Assessing acceptability and data completeness

Acceptability and data completeness will establish extent to which scale items are scored and total scores can be computed. Assessment of the completeness of item and scale-level data (i.e., missing or incomplete data for items and sample) including frequency of

endorsement will be completed. Score distributions including skew of scale scores and presence of floor and ceiling effects will be examined.[19,38,42]

Assessing scaling assumptions

Examination of scaling assumptions involves assessment of whether it is legitimate to group items into a scale to produce a scale score. Tests of scaling assumptions examine item-total correlations, mean scores and standard deviations.[19]

Assessing targeting

Targeting may be defined as "the extent to which the range of the variable measured by the scale matches the range of that variable in the study sample" (p.4).[43] In this study, targeting will be assessed using CTT methods by examining whether the LEQ-CI scale scores span the entire scale range, skewness, and whether floor and ceiling effects are low, defined as < 15% of the sample.[23]

Assessing internal consistency reliability

Assessment of internal consistency establishes the inter-relatedness among items and is an assessment of the unidimensionality of a scale or subscale.[44] Internal consistency will be assessed by calculating inter-item and item-total correlations and Cronbach's alpha.

- Inter-item correlation – Calculating inter-item correlations will provide an indication whether an item is part of a (sub)scale. Correlations should fall between 0.2 and 0.5. Items which have a correlation greater than 0.7 may be considered to measure the same thing, making one item a candidate for deletion.[23]
- Item-total correlation – Calculating item-total correlations will assess whether the LEQ-CI's items discriminate patients on the listening effort construct. Items that show an item-

total correlation of less than 0.3 will be considered as contributing little to the LEQ-CI in terms of discriminating between individuals with high versus low levels of listening effort.[23] These items will be considered candidates for deletion.

- Cronbach's alpha - Internal consistency will be calculated for each subscale of the LEQ-CI by calculating Cronbach's alpha.[19,45–47] Alpha values $\geq$ 0.70 and $\leq$ 0.95 will be considered good evidence of internal consistency.[19]

Phase 2: Establishing construct validity

Construct validity may be defined as the extent to which the scores of an instrument are a valid measure of the latent construct.[23] Construct validity of the refined LEQ-CI will be assessed by applying criteria specified by the COSMIN group. Construct validity may be assessed by testing a priori hypotheses about the relationship between the instrument and other measures, as well as the expected differences between the scores attained by different sub-groups of the target population based on the assumption that the LEQ-CI validly measures the target construct (i.e., listening effort). To establish the construct validity of an instrument, Mokkink et al. recommend at least 75% of the stated hypotheses are endorsed. [22]

Assessing convergent validity

Concurrent construct validity will be assessed by examining the correlation between scores on the LEQ-CI with the summed score on the three items considered to measure listening effort on the Speech, Spatial, and Qualities of Hearing questionnaire (SSQ).[48] As no validated measure of listening effort has been identified as a suitable comparator PROM these items were selected to assess construct validity as the SSQ has good evidence of being a well-validated instrument across multiple studies.[18] We hypothesise that a strong positive

correlation > 0.50 will be observed between measures. We further predict a moderate positive correlation (0.30 – 0.50) between the LEQ-CI and SSQ total score as LE may be considered to be a component of hearing disability, the construct measured by the SSQ.[5]

Assessing discriminant validity

Discriminant (i.e., divergent) validity is an assessment of a measure's ability to discriminate between dissimilar constructs.[19] It will be assessed by the examining the correlation between scores on the LEQ-CI and the Fatigue Assessment Scale (FAS),[49] a measure of fatigue used in other studies investigating LE and fatigue in individuals with HL.[50] As LE and fatigue are similar but unrelated constructs we anticipate a moderate positive correlation between 0.30 – 0.50.

Further assessment of discriminant validity will be undertaken by examining the correlation between scores on the LEQ-CI and the Nijmegen Cochlear Implant Questionnaire (NCIQ),[51] a measure of quality of life in CI patients. A small positive correlation of between 0.30 – 0.50 is anticipated as these measures may be considered to assess similar, but unrelated constructs. [22]

ETHICS AND DISEMINATION

This study has received ethical approval from the NHS Research Ethics Committee (REC): Newcastle and North Tyneside 2 (Ref: 18/NE/0320). Study findings will be disseminated through publication in peer-reviewed journals, conference presentations, and in the lead author's (SEH) doctoral dissertation.

Author contributions: SEH, HAH, AW and FR conceived the study. SH wrote the protocol and this manuscript. HAH, AW, FR, IB, CMM provided critical review of the protocol and this manuscript. All authors read and approved the final manuscript.

Funding: This project is supported by an Abertawe Bro Morgannwg University Health Board Pathway to Portfolio grant (R&D Pathway June002) awarded to Mrs. Sarah Hughes.

Competing interests: The authors have no competing interests to declare.

Provenance and peer review: Not commissioned; externally peer reviewed.

REFERENCES

1 Hutchings HA, Alrubaiy L. Patient-reported outcome measures in routine clinical care : The PROMise of a better future? *Dig Dis Sci* Published Online First: 2017. doi:10.1007/s10620-017-4658-z

2 Ousey K, Cook L. Understanding patient reported outcome measures (PROMs). *Br J Community Nurs* 2011;16:80–2.

3 Meadows KA. Patient-reported outcome measures: An overview. *Br J Community Nurs* 2011;16(3):146–51.

4 Devlin NJ, Appleby J, Buxton M, et al. Getting the Most Out of PROMs: Putting Health Outcomes at the Heart of NHS Decision-Making.London:The King's Fund 2010.

5 Noble W. Self-Assessment of Hearing. 2nd ed. San Diego:Plural Publishing 2013.

6 McGarrigle R, Munro K, Dawes P, et al. Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *Int J Audiol* 2014;53:433-40.

7 Mackersie CL, MacPhee IX, Heldt EW. Effects of hearing loss on heart rate variability and skin conductance measured during sentence recognition in noise. *Ear Hear* 2015;36:145-54.

8 Zekveld AA, George ELJ, Houtgast T, et al. Cognitive abilities relate to self-reported hearing disability. *J Speech Lang Hear Res* 2013;56:1364–72.

9 Ramage-Morin PL. Hearing difficulties and feelings of social isolation among Canadians aged 45 or older. *Health Rep* 2016;27:3–12.

10 Nachtegaal J, Kuik DJ, Anema JR, et al. Hearing status, need for recovery after work, and psychosocial work characteristics: results from an internet-based national survey on hearing. *Int J Audiol* 2009;48:684–91.

11 Miles K, McMahon C, Boisvert I, et al. Objective assessment of listening effort: Coregistration of pupillometry and EEG. *Trends Hear* 2017;21:233121651770639. doi:10.1177/2331216517706396

12 Pichora-Fuller MK, Kramer SE, Eckert MA, et al. Hearing impairment and cognitive energy: The Framework for Understanding Effortful Listening (FUEL). *Ear Hear* 2016;37(Suppl1):5S–27S.

13 Zekveld AA, Heslenfeld DJ, Johnsrude IS, et al. The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *Neuroimage* 2014;101:76–86.

14 Winn MB, Edwards JR, Litovsky RY. The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear* 2015;36:e153–65. doi:10.1097/AUD.0000000000000145

15 Bernarding C, Strauss DJ, Hannemann R, et al. Quantification of listening effort correlates in the oscillatory EEG activity: A feasibility study. *Conf Proc IEEE Eng Med Biol Soc* 2012;2012:4615–8.

16 McMahon CM, Boisvert I, de Lissa P, et al. Monitoring alpha oscillations and pupil dilation across a performance-intensity function. *Front Psychol* 2016;7:745. doi:10.3389/fpsyg.2016.00745

17 Gagné J, Besser J, Lemke U. Behavioral assessment of listening effort using a dual-task paradigm. *Trends Hear* 2017;21:1–25. doi:10.1177/2331216516687287

18 Hughes SE, Hutchings HA, Peters S, et al. A systematic review and narrative synthesis of the measurement properties of patient-reported outcome measures (PROMs) used to assess listening effort in hearing loss. *British Society of Audiology Conference 2017*. Harrogate: 2017.

19 Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. 4th ed. Oxford University Press 2008.

20 Hughes SE, Hutchings HA, Rapport FL, et al. Social connectedness and perceived listening effort in adult cochlear implant users: A Grounded Theory to establish content validity for a new patient-reported outcome measure. *Ear Hear* 2018;39:922–34.

21 Hughes SE, Rapport FL, Boisvert I, et al. Patient-reported outcome measures (PROMs) for assessing perceived listening effort in hearing loss: Protocol for a systematic review. *BMJ Open* 2017;7:e014995. doi:10.1136/bmjopen-2016-014995

22 Mokkink LB, de Vet HCW, Prinsen DL, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018; 27:1171-1179.

23 de Vet H, Terwee CB, Mokkink LB, et al. Measurement in Medicine: A Practical

Guide. Cambridge:Cambridge University Press 2011.

24 Linacre JM. Sample size and item calibration or person measure stability. *Rasch Meas Trans* 1994;7:328.

25 Hobart JC, Cano SJ, Warner TT, *et al.* What sample sizes for reliability and validity studies in neurology? *J Neurol* 2012;259:2681–94.

26 National Institute for Health and Clinical Excellence (NICE). NICE technology appraisal guidance 166: Cochlear implants for children and adults with severe to profound deafness. 2009. www.guidance.nice.org.uk/TA166/pdf/English (accessed 18 May 2018).

27 Boone WJ. Rasch analysis for instrument development: Why, when, and how? *CBE Life Sci Educ* 2016;15:pii-rm4.

28 Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences 3rd Ed. New York: Routledge 2015.

29 Petrillo J, Cano SJ, McLeod LD, et al. Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value Health* 2015;18:25–34.

30 Ball S, Vickery J, Hobart J, et al. The Cannabinoid Use in Progressive Inflammatory brain Disease (CUPID) trial: A randomised double-blind placebo-controlled parallel-group multicentre trial and economic evaluation of cannabinoids to slow progression in multiple sclerosis. *Health Technol Assess* 2015;19:1–187. doi:10.3310/hta19120

31 Tadic V, Cooper A, Cumberland P, et al. Measuring the quality of life of visually impaired children: First stage psychometric evaluation of the novel VQoL_CYP instrument. *PLoS One* 2016;11:e0146225.

32 Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol Assess* 1995;7:286–99.

33 Wright B, Linacre JM. Reasonable mean-square fit vales. *Rasch Anal Trans* 1994;8: 370.

34 Zumbo B. A handbook on the theory and methods of differential item functioning (DIF). *Ottawa Natl Def Hqrs* 1999:1–57.

35 Teresi JA, Ramirez M, Jones RN, *et al.* Modifying measures based on differential item functioning (DIF) impact analysis. *J Aging Health* 2012;24:1044–76.

36 Johnston M, Dixon D, Hart J, et al. Discriminant content validity: A quantitative methodology for assessing content of theory-based measures, with illustrative applications. *Br J Health Psychol* 2014;19:240–57.

37 Linacre JM. Category disordering (disordered categories) vs. threshold disordering (disordered thresholds). *Rasch Meas Trans* 1999;13:675.

38 Gorecki CA. *The development and validation of a patient-reported outcome measure of health-related qwuality of life for patients with pressure iulcers: PUQOL Project* University of Leeds 2011.

39 Tennant A, Conaghan, PG. The measurement model model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthiritis Rheum* 2007;57;1358-62.

40 Tractenberg RE. Classical and modern measurement theories, patient reports and clinical outcomes. *Comtemp Clin Trials* 2010;31:1–3.

41 da Rocha NS, Chachamovich E, de Almeida Fleck MP, et al. An introduction to Rasch analysis for psychiatric practice and research. *J Psychiatr Res* 2013;47:141–8.

42 McHorney CA, Ware, Jr. JE, Lu FR, et al. The MOS 36-Item Short-Form Health Survey (SF-36): III . Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;32:40–66.

43 Gorecki C, Brown JM, Cano S, et al. Development and validation of a new patient-

reported outcome measure for patients with pressure ulcers: The PU-QOL instrument. *Health Qual Life Outcomes* 2013;11:1.

44 Mokkink LB, Terwee CB, Patrick DL, et al. COSMIN checklist manual. Amsterdam: COSMIN 2012.

45 Streiner DL. Starting at the beginning: An introduction to coefficient alpha and internal consistency. *J Pers Assess* 2003;80:99–103.

46 Streiner DL. Being inconsistent about consistency: When coefficient apha does and doesn't matter. *J Pers Assess* 2003;80:217–22.

47 Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147-1157.

48 Gatehouse S, Noble W. The Speech, Spatial and Qualities of Hearing scale (SSQ). *Int J Audiol* 2004;43:85–99.

49 Michielsen HJ, De Vries J, Van Heck GL, et al. Examination of the dimensionality of fatigue. *Eur J Psychol Assess* 2004;20:39–48.

50 Alhanbali S, Dawes P, Lloyd S, et al. Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear Hear* 2017;38:e39–48.

51 Hinderink JB, Krabbe PF, Van Den Broek P. Development and application of a health-related quality-of-life instrument for adults with cochlear implants: The Nijmegen cochlear implant questionnaire. *Otolaryngol Head Neck Surg* 2000;123:756-65.

BMJ Open

Study protocol for the validation of a new patient-reported outcome measure (PROM) of listening effort in cochlear implantation: the Listening Effort Questionnaire-Cochlear Implant (LEQ-CI)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-028881.R1
Article Type:	Protocol
Date Submitted by the Author:	14-May-2019
Complete List of Authors:	Hughes, Sarah; Swansea University, Swansea University Medical School; South Wales Cochlear Implant Programme Rapport, Frances ; Macquarie University, Australian Institute of Health Innovation Watkins, Alan; Swansea University, Swansea University Medical School Boisvert, Isabelle; Macquarie University, Department of Linguistics (Audiology Section); Australian Hearing Hub, The HEARing CRC McMahon, Catherine; The HEARing CRC; Macquarie University, Department of Linguistics (Audiology Section) Hutchings, Hayley; Swansea University, Swansea University Medical School
Primary Subject Heading:	Ear, nose and throat/otolaryngology
Secondary Subject Heading:	Health services research
Keywords:	hearing loss, cochlear implant, patient-reported outcome measure, validation, listening effort, PROM

Study protocol for the validation of a new patient-reported outcome measure (PROM) of listening effort in cochlear implantation: the Listening Effort Questionnaire-Cochlear Implant (LEQ-CI)

Sarah E. Hughes[1,2]

Frances Rapport[3,6]

Alan Watkins[2]

Isabelle Boisvert[3,4,5]

Catherine M. McMahon[3,4,5]

Hayley A. Hutchings[2]

[1]South Wales Cochlear Implant Programme, Bridgend, Wales, United Kingdom

[2]Patient and Population Health and Informatics, Swansea University Medical School, Swansea, Wales, United Kingdom

[3]Centre for Implementation of Hearing Research, Macquarie University, Sydney, Australia

[4]Department of Linguistics (Audiology Section), Macquarie University, Sydney, Australia

[5]The HEARing Cooperative Research Centre, Australia

[6]Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

Corresponding Author:

Sarah E. Hughes

South Wales Cochlear Implant Programme

Princess of Wales Hospital

Coity Road, Bridgend

Wales, United Kingdom CF31 5EU

E-mail: sarah.hughes@wales.nhs.uk

Phone: +44 1656 752192

Fax: +44 1656 752192

Word count (excluding title page, abstract, references, figures and tables): 3740

Key words: hearing loss, cochlear implant, listening effort, patient-reported outcome measure, validation, PROM

ABSTRACT

Introduction: Listening effort may be defined as the cognitive resources needed to understand an auditory message. A sustained requirement for listening effort is known to have a negative impact on individuals' sense of social connectedness, well-being, and quality of life. A number of hearing-specific patient-reported outcome measures (PROMs) exist currently; however, none adequately assess listening effort as it is experienced in the listening situations of everyday life. The Listening Effort Questionnaire – Cochlear Implant (LEQ-CI) is a new, disease-specific PROM designed to assess perceived listening effort as experienced by adult CI patients. It is the aim of this study to conduct the first psychometric evaluation of the LEQ-CI's measurement properties.

Methods and analysis: This study is a phased, prospective, multi-site validation study in a UK population of adults with severe-profound sensorineural hearing loss (SNHL) who meet local candidacy criteria for CI. In Phase 1, 250 CI patients from four National Health Service (NHS) CI centres will self-complete a paper version of the LEQ-CI. Factor analysis will establish unidimensionality and Rasch analysis (RA) will evaluate item fit, differential item functioning (DIF), response scale ordering, targeting of persons and items, and reliability. Classical Test Theory (CTT) methods will assess acceptability/data completeness, scaling assumptions, targeting, and internal consistency reliability. Phase 1 results will inform refinements to the LEQ-CI. In Phase 2, a new sample of adult CI patients (n = 100) will self-complete the refined LEQ-CI, the Speech, Spatial and Qualities of Hearing Scale (SSQ), the Nijmegen Cochlear Implant Questionnaire (NCIQ) and the Fatigue Assessment Scale (FAS) to assess construct validity.

Ethics and dissemination: This study was approved by the Abertawe Bro Morgannwg University (ABMU) Health Board/Swansea University Joint Study Review Committee (JSRC) and the Newcastle and North Tyneside 2 Research Ethics Committee (REC), Ref: 18/NE/0320.

Dissemination will be in high-quality journals, conference presentations, and SEH's doctoral dissertation. (300 words)

ARTICLE SUMMARY

Strengths and limitations

- The LEQ-CI is the first PROM developed specifically to assess perceived listening effort in cochlear implant candidates and recipients.
- The proposed study conforms to international consensus standards on best practice of studies of instrument development and validation - the COnsensus-based Standards for the selection of health-status Measurement INstruments (COSMIN).
- The use of CTT and RA will enable a robust initial assessment of the LEQ-CI's measurement characteristics at both item and scale level.
- The conceptual framework underpinning the LEQ-CI is based on an explanatory model developed from current theoretical frameworks and the patient perspective. Assessment of the LEQ-CI's measurement properties will provide early evidence of the validity of the proposed model.
- Instrument validation is an iterative process to build a body of evidence relating to the quality of an instrument's measurement properties. Further studies that assess the measurement characteristics of LEQ-CI will be required.

INTRODUCTION

Hearing loss is a top ten burden of disease. It affects approximately one in every six people in the UK population and the economic burden is estimated to be over £30 billion annually.[1] Management of hearing loss is typically focussed on the provision of hearing technologies such as hearing aids or cochlear implants. However, even with appropriate provision of devices, individuals continue to report a sustained requirement for listening effort.[2] Listening effort may be defined as the mental exertion (the attentional and cognitive resources) needed to understand an auditory signal.[3] A sustained requirement for high listening effort is known to impact on the everyday listening activities of adults with hearing loss with negative implications for their social functioning, work recovery, social connectedness, well-being and quality of life.[4–7]

In the context of audiological clinical practice, many current routine assessments are capable of providing insight into audibility of the acoustic signal but are unable to supply information relating to the underlying processes and mechanisms, such as listening effort, that inform the measured performance. In the era of person-centred care, well-validated measures that assess these underlying factors are needed if hearing healthcare professionals are to adopt a more holistic approach to the management of hearing loss. Validated self-report instruments such as patient-reported outcome measures (PROMs) have the potential to be viable clinical measures of an individual's listening effort in everyday listening situations.

PROMs are self-report tools that assess an individual's perception of their disease severity, symptoms and functioning, quality of life or well-being.[8–10] PROMs are being used increasingly in routine clinical practice and are already well-established in the field of audiology.[11,12] PROMs enable clinicians to gain insight into the patient's perspective of their condition and the treatment they receive. Importantly, PROMs provide insight into those aspects of a disease or condition that are not observable, but rather, are knowable only to the

patient themselves. PROMs offer a complementary method to current behavioural (e.g., dual task paradigms) and physiological measures (e.g., pupilometry, fMRI, electroencephalography) of listening effort. There is a growing body of research to suggest that listening effort is a multidimensional construct and that these different measures may evaluate different aspects of this phenomenon.[4,13–16] Using factor analysis, Alhanbali et al. have shown that hearing level, SNR, dual-task paradigms, pupilometry and EEG (i.e., alpha power during speech recognition and retention) and self-reported effort tap into different underlying dimensions of listening effort.[14] Reflecting on this work, it may be argued that PROMs, as a measure of self-reported effort, have the potential to assess a dimension of listening effort that is not captured by current behavioural and physiological measures.

Several hearing-specific PROMs have been developed that include items considered to measure listening effort. A systematic review by the authors identified two PROMs that measured listening effort and cognitive effort in listening respectively.[17,18] Several PROMs assessing listening effort at either the item or subscale level (e.g., SSQ, (A)PHAB, CPHI) were also identified.[19–22] Overall, the review findings found limited evidence of these PROMs' psychometric measurement properties. The SSQ was identified as the current best candidate for use as a listening effort PROM based on the extent and quality of its validation when assessed against the COSMIN criteria.[23] However, one drawback of the SSQ as a measure of listening effort is a high response burden with only 6% of its items measuring listening effort. Notably, all of the PROMs identified in this systematic review were developed prior to publication of the theoretical frameworks and treatises that inform current conceptualisations of listening effort including the role of motivation on effort expenditure.[2,24–26] Lack of congruence between these instruments and current frameworks is a limitation of the content validity of existing PROMs. It is unlikely these instruments capture fully the conceptualisation of listening effort as presented in these recently published models. As such, there is growing

support in the literature for a new PROM that comprehensively measures self-reported listening effort in hearing loss as it is conceptualised currently.[14,26] To address this situation, the Listening Effort Questionnaire – Cochlear Implant (LEQ-CI) has been developed. The LEQ-CI is a new hearing-specific PROM measuring perceived listening effort in adults who receive cochlear implants.

AIMS

To have confidence that a PROM is providing meaningful information, psychometric evaluation of its measurement properties must be undertaken to satisfy rigorous criteria.[23,27] This includes assessment of an instrument's validity (i.e., to what extent does the instrument measure the construct it purports to measure), its reliability (i.e., the degree to which measurement is free from error) and its responsiveness (i.e., the ability of an outcome measure to detect change over time in the construct to be measured).[28] There are several measurement properties that require assessment and each property needs its own type of study to assess it. The process of psychometric validation is iterative and represents an accumulation of evidence over time from multiple studies.[29]

The aim of this study is to conduct the first psychometric validation of the LEQ-CI in accordance with the internationally recognised COSMIN guidelines.[28,30] Building on previous work undertaken by the authors to establish the LEQ-CI's content validity,[2,31] the current study represents a further step towards the provision of a robust self-report measure of perceived listening effort for use in research and clinical practice.

OBJECTIVES

- To refine the items, response categories, and scale structure of the new LEQ-CI using Rasch Measurement Theory in an English-speaking sample of adult cochlear implant candidates and recipients in the UK.

- To assess acceptability, scaling assumptions, targeting, and reliability using CTT methods.
- To assess the construct validity of the refined LEQ-CI in the population of adults with severe-profound, post-lingual HL who are CI candidates or recipients.

METHODS AND ANALYSIS

Study Setting and Patient Involvement

This prospective study is a UK-based multi-phase study to validate the measurement properties of a new PROM, the LEQ-CI. The planned study will take place over a 12-month period and has been co-produced by the study team with input from two lay members, both CI recipients, from the study's Research Management Group. Lay members reviewed and provided feedback on study design, participant documents and iterations of the LEQ-CI.

Development of the LEQ-CI

The LEQ-CI is a hearing-specific PROM measuring listening effort in adult cochlear implant candidates and recipients. It is comprised of 29 items across four domains. Five-point or seven-point Likert scales with absolute anchors and labelled categories ensure a broad range of response options. Item responses are summed to produce a simple total score. The LEQ-CI's conceptual framework, presented in Figure 1, was developed from a mixed-methods qualitative study involving focus groups and a postal survey.[2] An item bank was constructed that included new items and items harvested from extant PROMs considered to measure listening effort or associated constructs.[17,18] Exemplar items are presented in Figure 2. Preliminary testing was completed to identify and rectify problems with items and response scales prior to undertaking psychometric evaluation.[31] Preliminary testing involved the use of multiple datasets to assess reliability.[32] Item quality was estimated using the on-line Survey Quality Predictor system (SQP 2.1, http://sqp.upf.edu/). An expert review panel of academics,

researchers, and clinicians (n = 7) and a series of cognitive interviews with a purposive sample of cochlear implant candidates and recipients (n = 12) were completed to elicit feedback on the relevance, clarity and acceptability of the LEQ-CI.

Sample size

The study sample will be representative of the population of adults with acquired, post-lingual severe-to-profound sensorineural hearing loss (SNHL) referred for cochlear implantation in the UK. A total study sample of 350 participants will be recruited. In Phase 1, a cohort of 250 participants will be recruited from four National Health Service (NHS) cochlear implant centres. To minimise burden on implant centre staff and to ensure representation from different regions of the UK, each centre will send questionnaire packs to 125 cochlear implant candidates or recipients who meet the study inclusion criteria (n = 500). If necessary, additional participants will be recruited until such time as 250 completed LEQ-CI forms with no missing data are returned. There are no general criteria for determination of sample sizes in studies of PROM validation and sample sizes are, in part, dependent on the psychometric characteristics being assessed.[23,27] Mokkink et al. recommend greater than 200 respondents when undertaking RA and seven times the number of items for purposes of undertaking assessment of unidimensionality.[23] Linacre recommends a minimum sample size of 250 respondents for definitive item calibration using RA.[33] The LEQ-CI is comprised of 29 items that have been selected to minimise respondent burden whilst allowing for adequate sampling of relevant constructs associated with listening effort. Therefore, a minimum sample of 250 participants is considered sufficient for undertaking both assessment of unidimensionality and RA of the LEQ-CI. In Phase 2, a new cohort of 100 participants fulfilling the same eligibility criteria will be recruited from two cochlear implant centres. Each centre will recruit 125 participants initially. If necessary, further participants will be recruited until the required sample size is

achieved. Hobart et al. recommend greater than 80 participants for assessment of construct validity.[34]

Recruitment and data collection

The participant eligibility criteria are the same for both phases of the study and are presented in Table 1.

Table 1. Study eligibility criteria for recruitment of participants

Inclusion Criteria	Exclusion Criteria
• Adults (persons ≥ 18 years of age). • Post-lingual severe-profound SNHL • A candidate for cochlear implantation according to UK criteria specified by the National Institute of Health and Care Excellence (NICE, 2009) or cochlear implant recipient.[35] • Proficient readers/writers of English. • Capacity to give informed consent. • No additional medical conditions precluding the participant's ability to self-complete the questionnaires	• Pre-lingual severe-to-profound SNHL (i.e., when the onset of the hearing loss can reasonably be estimated to have occurred before age 3, in both ears) and the individual's primary mode of communication is manual (e.g., British Sign Language).

In Phases 1 and 2, participants meeting the study inclusion criteria will be sent an invitation letter, an information sheet describing the study in detail, the LEQ-CI, a demographic questionnaire, and comparator questionnaires (Phase 2 participants only). A reply-paid envelope for the return of the completed questionnaires will be provided. Informed consent is

presumed if the questionnaires are completed and returned to the study team. To maintain participant anonymity, eligibility screening and the study documents will be mailed to prospective participants by a member of the clinical team at each participating cochlear implant centre. To maintain participant anonymity, each questionnaire pack will be coded with a unique identifier. No personal identifiable information will be retained by the study team.

Statistical Analysis

There are two schools of psychometric measurement theory dominate the field of PROM development.[29,36] Traditional psychometric analyses (e.g., Cronbach's alpha as a measure of internal consistency reliability) are underpinned by CTT. CTT seeks to evaluate reliability and validity of a scale and has been the dominant approach used in the development and validation of outcome measures.[37] However, modern measurement techniques such as RA are increasingly being reported alongside traditional analyses in studies of PROM development and validation (e.g., [38,39]).

CTT is based on the assumption that every observed score is a function of an individual's true score and random error.[40] The assumptions underpinning CTT differ from those underpinning the Rasch model. It has been argued that CTT cannot be adequately be tested as it is based on definitions rather than assumptions which can be proven true or false. This is in contrast to modern measurement theory (i.e., RA) which can generate assumptions that can be proven true or false.[41] Whereas CTT methods focus on the total score of a measure, RA enables instrument developers to focus more specifically on the characteristics of individual items.[42] For example, RA, unlike CTT methods, can be used to establish whether an item's response scale is functioning as expected and, if not, suggest improvements.

The Rasch model allows for ordering persons (i.e., patients) according to the amount of the latent target construct (i.e., listening effort) they possess and for ordering items that

measure the target construct according to their difficulty.[36] This method allows non-linear (i.e., ordinal) raw data to be converted to a linear (i.e., interval) scale, which can then be evaluated through the use of parametric statistical tests.[43] By contrast, CTT methods yield measures that produce ordinal rather than interval level data. This has implications for the interpretation of test scores as difference scores and changes scores are most meaningful when interval level of measurement is used.[36,41]

A further limitation of CTT is that the performance of a test is dependent on the sample in which that test is assessed.[41] This renders its psychometric properties (i.e., reliability and validity) dependent on the sample rather than characteristics of the test itself. By contrast, RA produces item and test statistics that are sample independent rendering the test valid across groups. Any discrepancies between the scale data and the Rasch model requirements are indicative of anomalies in the scale as a measurement instrument. These discrepancies provide diagnostic information that serves as a basis for understanding and empirical improvement of the instrument at both item and scale-level.[44]

Despite these limitations, CTT methods continue to be widely used in studies of instrument validation and are included in the COSMIN standards.[28,41] Indeed, some properties (e.g., acceptability, scaling assumptions) can only be evaluated using CTT methods.[37] For these reasons, this study will use both CTT and RA in a complementary fashion to ensure rigorous validation of the LEQ-CI at both item and scale level.

Study data will be managed using the online clinical data management programme, REDCap (Version 7.2.1, Vanderbilt University), licensed to the Swansea Trials Unit, Swansea University. RA will be used to evaluate the LEQ-CI's psychometric properties and refine the item and scale structure of the LEQ-CI using Winsteps (Version 4.1.0) software. Psychometric analyses applying CTT will be conducted using the Statistical Package for Social Sciences (SPSS), Version 22.0 licensed to Swansea University.

Phase 1: Item & Scale Refinement using Rasch Measurement Theory

Assessing unidimensionality

The Rasch measurement model assumes unidimensionality which is defined as the measurement of a single latent construct.[36,45] Therefore, prior to undertaking RA, factor analysis will be undertaken to assess the underlying structure of the LEQ-CI and establish the unidimensionality of its (sub)scales.[46]

Assessing item fit

In RA, item fit refers to the degree of mismatch between the pattern of actual observed responses and the Rasch modelled expectations. Specifically, it refers to the pattern for each item across persons investigated by examining item infit and outfit statistics.[36] Mean square standardized residuals (MNSQ) will be used to assess fit with MNSQ residuals within the 0.5–1.5 range considered acceptable for productive measurement. Mean square values less than 0.5 indicate overfit (i.e., the items are too predictable relative to the Rasch model), while mean square values greater than 1.5 are indicative of too much noise (randomness) relative to the Rasch model.[47]

Assessing differential item functioning (DIF)

DIF is an indication of the loss of invariance across subsamples of respondents. The presence of DIF will be an indicator of potential problems with an item since item and person measures on a unidimensional instrument should remain invariant (i.e., within error) across all appropriate measurement conditions.[36] DIF will be examined for key demographic variables such as age and sex. The standard threshold of > 1 logit will be used as an indicator of DIF.[45] The methodology proposed by Zumbo using logistic regression DIF tests for significance (i.e., Chi-squared two degrees of freedom test) and magnitude of DIF by computing the R-squared effect for both uniform and non-uniform DIF will be applied.[48] If items are found to have

DIF they will either be considered candidates for removal or examined for adjustment of DIF and re-evaluated, thus reflecting the iterative nature of instrument validation.[36,49]

Assessing response scale ordering

The response options of an instrument (i.e., number of categories and their definitions) are critical to its reliability and validity.[50] The Rasch model will enable us to show empirically how respondents use the LEQ-CI's rating scale informing future iterations of the LEQ-CI to ensure it yields high quality data.[36] Response category ordering will be assessed using Rasch probability curves and there will be an examination of the data for category disordering and threshold disordering.[51] These investigations will show whether the response options selected for the LEQ-CI are sufficient or should be collapsed to provide better coverage of the latent trait.

Assessing the targeting of persons and items

Targeting using RA explores whether the instrument has a distribution of items that matches the range of the respondents' latent trait. This will be done by examining the item-person threshold distribution map, which illustrates a relative position of "item difficulty" to "person ability".[45] The means and standard deviations of items and persons should match closely.[36]

Assessing reliability

The reliability of the LEQ-CI will be examined by observing the person separation index (PSI). The PSI is an estimate of the spread or separation of persons on the measured variable [36] and is considered to be a measure of internal consistency reliability.[52] It is a measure of the scale's ability to separate the study sample. A PSI > 0.7 will be considered an adequate measure of reliability.[53]

Phase 1: Psychometric evaluation using Classical Test Theory

Assessing acceptability and data completeness

Acceptability and data completeness will establish extent to which scale items are scored and total scores can be computed. Assessment of the completeness of item and scale-level data (i.e., missing or incomplete data for items and sample) including frequency of endorsement will be completed. Score distributions including skew of scale scores and presence of floor and ceiling effects will be examined.[29]

Assessing scaling assumptions

Examination of scaling assumptions involves assessment of whether it is legitimate to group items into a scale to produce a scale score. Tests of scaling assumptions examine item-total correlations, mean scores and standard deviations. When checking homogeneity of the LEQ-CI's scales, the heuristic that items should correlate with the total score above 0.20 will be applied. Item-total correlations will be calculated using the Pearson product-moment correlation.[29]

Assessing targeting

Targeting may be defined as "the extent to which the range of the variable measured by the scale matches the range of that variable in the study sample" (p.4).[54] Targeting will be assessed following item refinement using RA. CTT will be used determine whether the LEQ-CI scale scores span the entire scale range, skewness, and whether floor and ceiling effects are low, defined as < 15% of the sample.[27]

Assessing internal consistency reliability

Assessment of internal consistency establishes the inter-relatedness among items and is an assessment of the unidimensionality of a scale or subscale.[23] Internal consistency will be assessed by calculating inter-item and item-total correlations and Cronbach's alpha.

- Inter-item correlation – Calculating inter-item correlations will provide an indication whether an item is part of a (sub)scale. Correlations should fall between 0.2 and 0.5. Items

which have a correlation greater than 0.7 may be considered to measure the same thing, making one item a candidate for deletion.[27]

- Item-total correlation – Calculating item-total correlations will assess whether the LEQ-CI's items discriminate patients on the listening effort construct. Items that show an item-total correlation of less than 0.3 will be considered as contributing little to the LEQ-CI in terms of discriminating between individuals with high versus low levels of listening effort.[27] These items will be considered candidates for deletion.
- Cronbach's alpha - Internal consistency will be calculated for each subscale of the LEQ-CI by calculating Cronbach's alpha. Alpha values $\geq$ 0.70 and $\leq$ 0.95 will be considered good evidence of internal consistency.[29]

Phase 2: Establishing construct validity

Construct validity may be defined as the extent to which the scores of an instrument are a valid measure of the latent construct.[27] Construct validity of the refined LEQ-CI will be assessed by applying criteria specified by the COSMIN group. COSMIN guidance specifies that construct validity may be assessed by testing a priori hypotheses based on the literature and the experience of the study team.[23] Hypotheses are generated by the study team and founded on the assumption that the LEQ-CI validly measures the target construct (i.e., listening effort). These state the relationship between the instrument and other measures, as well as the expected differences between the scores attained by different sub-groups of the target population.. To establish the construct validity of an instrument, Mokkink et al. recommend at least 75% of the stated hypotheses are endorsed.[23]

Assessing convergent validity

Concurrent construct validity will be assessed by examining the correlation between scores on the LEQ-CI with the summed score on the three items considered to measure listening

effort on the Speech, Spatial, and Qualities of Hearing questionnaire (SSQ). As no validated measure of listening effort has been identified as a suitable comparator PROM these items were selected to assess construct validity as the SSQ has good evidence of being a well-validated instrument across multiple studies.[17] As the LEQ-CI and the SSQ are measuring the same construct, we hypothesise that a strong positive correlation > 0.50 will be observed between measures as suggested by Mokkink et al.[23] We further predict a moderate positive correlation (0.30 – 0.50) between the LEQ-CI and SSQ total score as LE may be considered to be a component of hearing disability, the construct measured by the SSQ.[11]

Assessing discriminant validity

Discriminant (i.e., divergent) validity is an assessment of a measure's ability to discriminate between dissimilar constructs.[29] It will be assessed by the examining the correlation between scores on the LEQ-CI and the Fatigue Assessment Scale (FAS),[55] a measure of fatigue used in other studies investigating LE and fatigue in individuals with HL.[56] As LE and fatigue are similar but unrelated constructs [16] we anticipate a moderate positive correlation between 0.30 – 0.50.

Further assessment of discriminant validity will be undertaken by examining the correlation between scores on the LEQ-CI and the Nijmegen Cochlear Implant Questionnaire (NCIQ),[57] a measure of quality of life in CI patients. A small positive correlation of between 0.30 – 0.50 is anticipated as these measures may be considered to assess similar, but unrelated constructs.[23]

ETHICS AND DISSEMINATION

This study has received ethical approval from the NHS Research Ethics Committee (REC): Newcastle and North Tyneside 2 (Ref: 18/NE/0320). Study findings will be

disseminated through publication in peer-reviewed journals, conference presentations, and in the lead author's (SEH) doctoral dissertation.

Author contributions: SEH, HAH, AW and FR conceived the study. SEH wrote the protocol and this manuscript. HAH, AW, FR, IB, CMM provided critical review of the protocol and this manuscript. All authors read and approved the final manuscript.

Funding: This project is supported by an Abertawe Bro Morgannwg University Health Board Pathway to Portfolio grant (R&D Pathway June002) awarded to Mrs. Sarah Hughes.

Competing interests: The authors have no competing interests to declare.

Provenance and peer review: Not commissioned; externally peer reviewed.

REFERENCES

1 Wilson BS, Tucci DL, Merson MH, *et al.* Global hearing health care: new findings and perspectives. *Lancet* 2017;**390**:2503–15.

2 Hughes SE, Hutchings HA, Rapport FL, *et al.* Social connectedness and perceived listening effort in adult cochlear implant users: a Grounded Theory to establish content validity for a new patient reported outcome measure. *Ear Hear* 2018;**39**:922-34.

3 McGarrigle R, Munro K, Dawes, *et al.* Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper'. *Int J Audiol* 2014;**53**:433-40.

4 Mackersie CL, MacPhee IX, Heldt EW. Effects of hearing loss on heart rate variability and skin conductance measured during sentence recognition in noise. *Ear Hear* 2015;**36**:145-54..

5 Zekveld AA, George ELJ, Houtgast T, *et al.* Cognitive abilities relate to self-reported hearing disability. *J Speech Lang Hear Res* 2013;**56**:1364–72.

6 Ramage-Morin PL. Hearing difficulties and feelings of social isolation among Canadians aged 45 or older. 2016;**27**:3-12.http://www.statcan.gc.ca/pub/82-003-x/2016011/article/14671-eng.pdf (accessed 23 Apr 2017).

7 Nachtegaal J, Kuik DJ, Anema JR, *et al.* Hearing status, need for recovery after work, and psychosocial work characteristics: results from an internet-based national survey on hearing. *Int J Audiol* 2009;**48**:684–91.

8 Hutchings HA, Alrubaiy L. Patient-Reported Outcome Measures in Routine Clinical Care : The PROMise of a Better Future? *Dig Dis Sci*.2017;**62**:1841-43.

9 Ousey K, Cook L. Understanding patient reported outcome measures (PROMs). *Br J Community Nurs* 2011;**16**:80–2.

10 Meadows KA. Patient-reported outcome measures: an overview. *Br J Community Nurs* 2011;**16**:146–51.

11 Noble W. Self-Assessment of Hearing. 2nd ed. San Diego:Plural Publishing 2013.

12 Devlin NJ, Appleby J. Getting the most out of PROMs: putting health outcomes at the heart of NHS decision-mkaing. London: 2010. http://www.kingsfund.org.uk/sites/files/kf/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf (accessed 16 Feb 2014).

13 Hornsby BWY. The effects of hearing aid use on listening effort and mental fatigue associated with sustained speech processing demands. *Ear Hear* 2013;**34**:523–34.

14 Alhanbali S, Dawes P, Millman RE, *et al.* Measures of listening effort are

multidimensional. *Ear Hear* 2019;e-pub ahead of print:1–14.

15 Miles K, McMahon C, Boisvert I, *et al.* Objective assessment of listening effort: coregistration of pupillometry and EEG. *Trends Hear* 2017;**21**:1-13.

16 Alhanbali S, Dawes P, Lloyd S, *et al.* Hearing handicap and speech recognition correlate with self-reported listening effort and fatigue. *Ear Hear* 2018;39:470-74.

17 Hughes SE, Hutchings HA, Dobbs TD, *et al.* A systematic review and narrative synthesis of the measurement properties of patient-reported outcome measures (PROMs) used to assess listening effort in hearing loss [poster presentation]. British Society of Audiology 2017;Harrogate.

18 Hughes SE, Rapport FL, Boisvert I, *et al.* Patient-reported outcome measures (PROMs) for assessing perceived listening effort in hearing loss: protocol for a systematic review. *BMJ Open* 2017 June 6;7(5):e014995.

19 Gatehouse S, Noble W. The speech, spatial and qualities of hearing scale (SSQ). *Int J Audiol* 2004;**43**:85–99.

20 Cox RM, Gilmore C, Alexander GC. Comparison of two questionnaires for patient-assessed hearing aid benefit. *J Am Acad Audiol* 1991;**2**:134–45.

21 Cox RM, Alexander GC. The Abbreviated Profile of Hearing Aid Benefit. *Ear Hear* 1995;**16**:176–86.

22 Demorest ME, Erdman SUEANN. Development of the commuication profile for the hearing impaired. *J Speech Hear Disord* 1987;**52**:129–43.

23 Mokkink LB, Prinsen CAC, Patrick DL, *et al.* COSMIN Methodology for Systematic Reviews of Patient - Reported Outcome Measures (PROMs) User Manual Version 1.0. 2018 February:1–78. www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf (accessed 4 Apr 2019)

24 Strauss DJ, Francis AL. Toward a taxonomic model of attention in effortful listening.

Cogn Affect Behav Neurosci 2017;**17**:809–25.

25 Peelle JE. Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear* 2018;39:204-14.

26 Pichora-Fuller MK, Kramer SE, Eckert MA, *et al.* Hearing impairment and cognitive energy. *Ear Hear* 2016;**37**:5S-27S.

27 de Vet H, Terwee CB, Mokkink LB, *et al.* Measurement in Medicine: A Practical Guide. Cambridge:Cambridge University Press 2011.

28 Mokkink LB, de Vet HCW, Prinsen CAC, *et al.* COSMIN Risk of bias checklist for sysmtematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;**27**:1171-79.

29 Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their Development and Use. 4th ed. Oxford University Press 2008.

30 Prinsen CAC, Mokkink LB, Bouter LM, *et al.* COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;**27**:1147–57.

31 Hughes SE, Boisvert I, Rapport F, *et al.* Measuring the experience of listeing effort in cochlear implant candidates and recipients: Development and pretesting of the LEQ-CI [poster presentation]. Southampton: British Cochlear Implant Group Academic Meeting 2019.

32 Maitland A, Presser S. How accurately do different evaluation methods predict the reliability of survey questions? *J Surv Stat Methodol* 2016;**4**:362-81.

33 Linacre JM. Sample size and item calibration or person measure stability. *Rasch Meas Trans* 1994;**7**:328.

34 Hobart JC, Cano SJ, Warner TT, *et al.* What sample sizes for reliability and validity studies in neurology? *J Neurol* 2012;**259**:2681–94.

35 National Institute for Health and Clinical Excellence (NICE). NICE technology

appraisal guidance 166: Cochlear implants for children and adults with severe to profound deafness. 2009. www.guidance.nice.org.uk/TA166/pdf/English (accessed Dec 2018).

36 Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences 3rd ed. London:Routledge 2015.

37 Petrillo J, Cano SJ, Mcleod LD, *et al.* Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value Heal* 2015;**18**:25–34.

38 Barker BA, Donovan NJ, Schubert AD, *et al.* Using Rasch analysis to examine the item-level psychometrics of the *Infant-Toddler Meaningful Auditory Integration Scales*. *Speech, Lang Hear* 2017;**20**:130–43.

39 Heffernan E, Maidment DW, Barry JG, *et al.* Refinement and validation of the Social Participation Restrictions Questionnaire: an application of Rasch analysis and traditional psychometric analysis techniques. *Ear Hear* 2018;40:328-39.

40 Tractenberg RE. Classical and modern measurement theories, patient reports and clinical outcomes. *Comtemp Clin Trials* 2010;**31**:1–3.

41 Cano SJ, Hobart JC, Cano S. The problem with health measurement. *Patient Prefer Adherence* 2011;**5**:279-90.

42 da Rocha, NS, Chachamovich E, de Almeida Fleck, MP *et al.* An introduction to Rasch analysis for psychiatric practice and research. *J Psychiatr Res* 2013;**47**:141–8.

43 Boone WJ. Rasch analysis for instrument development: why, when, and how? *CBE Life Sci Educ* 2016;**15**: rm4.

44 Ball S, Vickery J, Hobart J, *et al.* The Cannabinoid Use in Progressive Inflammatory brain Disease (CUPID) trial: a randomised double-blind placebo-controlled parallel-group multicentre trial and economic evaluation of cannabinoids to slow progression

in multiple sclerosis. *Health Technol Assess* 2015;**19**:1–187.

45 Tadic V, Tadi V, Cooper A, *et al.* Measuring the quality of life of visually impaired children: first stage psychometric evaluation of the novel VQoL-CYP. *PLoS One* 2016;**11**:e0146225.

46 Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol. Assess* 1995;**7**:286–99.

47 Wright, Linacre. Reasonable mean-square fit vales. *Rasch Meas Trans* 1994;**8**.www.rasch.org (accessed 20 Feb 2018).

48 Zumbo B. A handbook on the theory and methods of differential item functioning (DIF): logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores. *Directorate of Human Resources Research and Evaluation, Department of National Defense* 1999.www.researchgate.net/publication/236596822_A_handbook_on_the_theory_and_methods_of_differential_item_functioning_(DIF)_Logistic_regression_modeling_as_a_unitary_framework_for_binary_and_Likert-type_(ordinal)_item_scores/file/60b7d51830c07e4cbc.pdf (accessed Dec-2018).

49 Teresi JA, Ramirez M, Jones RN, *et al.* Modifying measures based on differential item functioning (DIF) impact analyses. *J Aging Health* 2012;**24**:1044–76.

50 Johnston M, Dixon D, Hart J, *et al.* Discriminant content validity: a quantitative methodology for assessing content of theory-based measures, with illustrative applications. *Br J Health Psychol* 2014;**19**:240–57.

51 Linacre JM. Category Disordering (disordered categories) vs. Threshold Disordering (disordered thresholds). *Rasch Meas Trans* 1999;**13**:675.

52 Gorecki CA. The development and validation of a patient-reported outcome measure of health-related quality of life for patients with pressure ulcers: PUQOL Project.

University of Leeds 2011.

53 Tennant, Conaghan. The measurement model model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care Res* 2007;**57**:1358-62

54 Gorecki C, Brown JM, Cano S, *et al.* Development and validation of a new patient-reported outcome measure for patients with pressure ulcers: the PU-QOL instrument. *Health Qual Life Outcomes* 2013;**11**:95.

55 Michielsen HJ, De Vries J, Van Heck GL, *et al.* Examination of the dimensionality of fatigue. *Eur J Psychol Assess* 2004;**20**:39–48.

56 Alhanbali S, Dawes P, Lloyd S, *et al.* Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear Hear* 2017;**38**:e39–48.

57 Hinderink JB, Krabbe PF, Van Den Broek P. Development and application of a health-related quality-of-life instrument for adults with cochlear implants: the Nijmegen Cochlear Implant Questionnaire. *Otolaryngol Head Neck Surg* 2000;**123**:756–65.

FIGURE 1

The conceptual frame work of the LEQ-CI

FIGURE 2

Example items from the LEQ-CI

Figure 1. The conceptual framework of the LEQ-CI

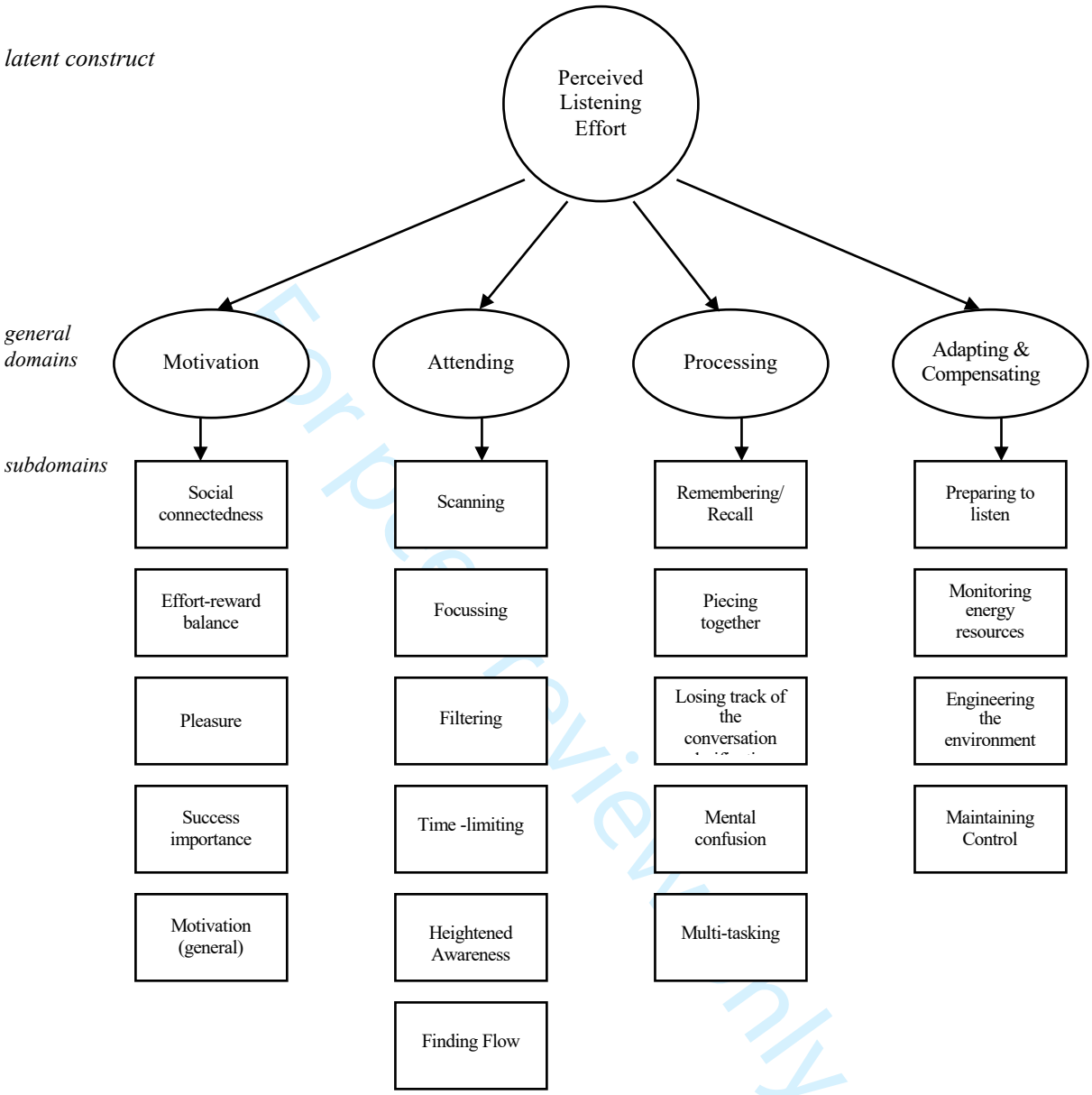

Figure 2: Example items from the LEQ-CI

Domain	**Item**	
Attending	E31	Many people with a hearing loss say it is difficult to keep listening for more than a few minutes at a time. In a **typical week**, are you able to listen for as long as is needed? *Please tick one box.* **Response Scale:** 7-point Likert scale (1 = never; 2 = rarely; 3 = occasionally; 4 = sometimes; 5 = frequently; 6 = usually; 7 = always)
Processing	E323	In a typical week, are you able to listen to someone talk while doing something else? *Please tick one box.* **Response Scale:** 7-point Likert scale (1 = never; 2 = rarely; 3 = occasionally; 4 = sometimes; 5 = frequently; 6 = usually; 7 = always)
Adapting & compensating	E105	In a **typical week**, on how many days do you run out of energy for listening before the end of the day? *Please tick one box.* **Response Scale:** 5-point Likert scale (1 = never; 2 = 1-2 days; 3 = 3-4 days; 4 = 5-6 days; 5 = Everyday)
Motivation	E326	In a **typical week**, does the effort of listening ever stop you from doing the things you want to do? *Please tick one box.* **Response Scale**: 7-point Likert scale (1 = never; 2 = rarely; 3 = occasionally; 4 = sometimes; 5 = frequently; 6 = usually; 7 = always)