

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)

Sarah E Hughes,^{1,2} Frances Rapport,³ Alan Watkins,¹ Isabelle Boisvert,^{4,5} Catherine M McMahon,^{4,5} Hayley A Hutchings¹

To cite: Hughes SE, Rapport F, Watkins A, *et al.* 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). *BMJ Open* 2019;**9**:e028881. doi:10.1136/bmjopen-2018-028881

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-028881>).

Received 31 December 2018
Revised 14 May 2019
Accepted 17 May 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Sarah E Hughes;
sarah.hughes@wales.nhs.uk

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, hearing-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 who meet local candidacy criteria for CI. In phase 1, 250 CI patients from four National Health Service CI centres will self-complete a paper version of the LEQ-CI. Factor analysis will establish unidimensionality and Rasch analysis will evaluate item fit, differential item functioning, response scale ordering, targeting of persons and items, and reliability. Classical test theory 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, the Nijmegen Cochlear Implant Questionnaire and the Fatigue Assessment Scale to assess construct validity.

Ethics and dissemination This study was approved by the Abertawe Bro Morgannwg University Health Board/Swansea University Joint Study Review Committee and the Newcastle and North Tyneside 2 Research Ethics Committee, Ref: 18/NE/0320. Dissemination will be in high-quality journals, conference presentations and SEH's doctoral dissertation.

Strengths and limitations of this study

- The Listening Effort Questionnaire-Cochlear Implant (LEQ-CI) is the first patient-reported outcome measure 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 classical test theory 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

Hearing loss is a top 10 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 focused on the provision of hearing technologies such as hearing aids or cochlear implants (CI). 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, these instruments 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 (eg, dual-task paradigms) and physiological measures (eg, pupillometry, functional MRI (fMRI), electroencephalography (EEG)) 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, signal to noise ratio (SNR), dual-task paradigms, pupillometry and EEG (ie, 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 (eg, Speech, Spatial and Qualities of Hearing Scale (SSQ), Profile of Hearing Aid Benefit (PHAB), Communication Profile for the Hearing Impaired (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 CONsensus-based STANDards for the selection of health-status Measurement INSTRUMENTs (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 that inform current conceptualisations of listening effort.^{2 24–26} Lack of congruence between these instruments and current theoretical frameworks is a limitation of the content validity of existing PROMs. It is unlikely these instruments capture fully listening effort as presented in these recently published models and, 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 CIs.

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 (ie, to what extent does the instrument measure the construct it purports to measure), its reliability (ie, the degree to which measurement is free from error) and its responsiveness (ie, 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 CI candidates and recipients in the UK.
- To assess acceptability, scaling assumptions, targeting and reliability using classical test theory (CTT) methods.

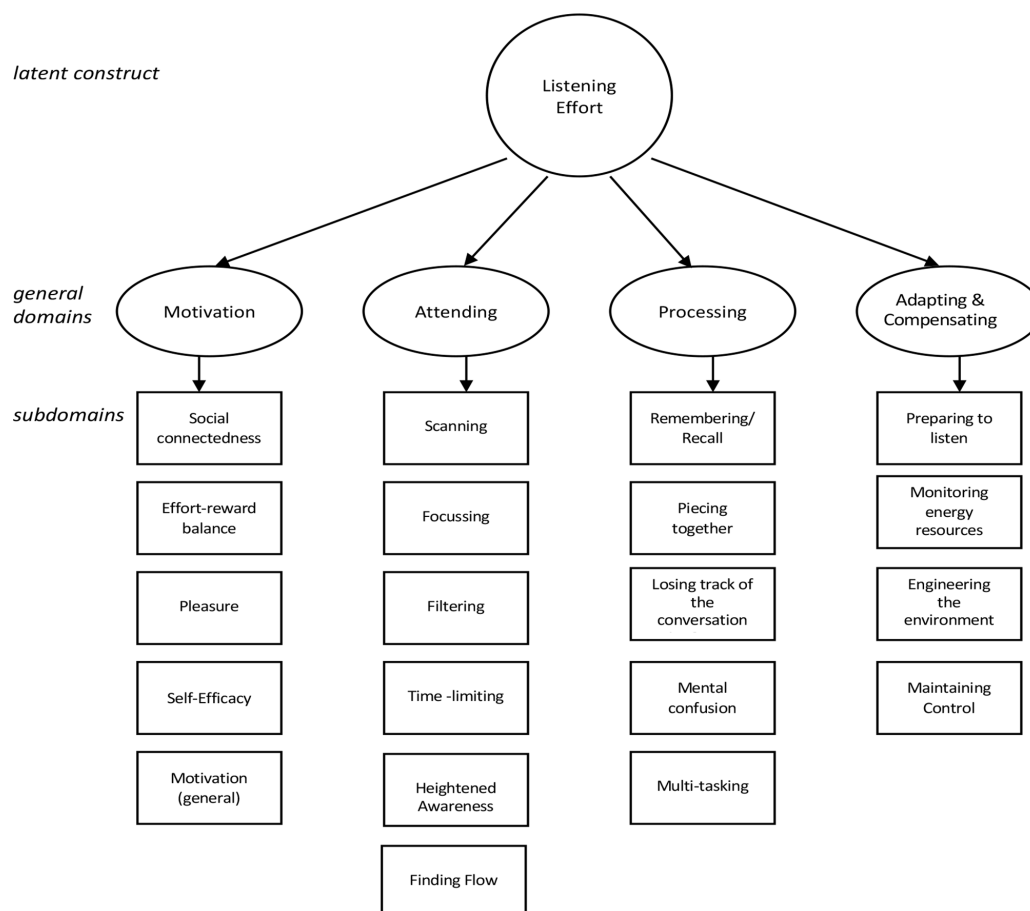


Figure 1 The conceptual framework of the Listening Effort Questionnaire-Cochlear Implant (LEQ-CI).

- To assess the construct validity of the refined LEQ-CI in the population of adults with severe-profound, postlingual HL who are CI candidates or recipients.

METHODS AND ANALYSIS

Study setting and patient involvement

This prospective study is a UK-based multiphase 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 coproduced 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 CI candidates and recipients. It is composed 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

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? <i>Please tick one box.</i> 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? <i>Please tick one box.</i> 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? <i>Please tick one box.</i> 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? <i>Please tick one box.</i> Response Scale: 7-point Likert scale (1 = never; 2 = rarely; 3 = occasionally; 4 = sometimes; 5 = frequently; 6 = usually; 7 = always)

Figure 2 Example items from the Listening Effort Questionnaire-Cochlear Implant (LEQ-CI).

to undertaking psychometric evaluation.³¹ Preliminary testing involved the use of multiple datasets to assess reliability.³² Item quality was estimated using the online (SQP V.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 CI 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, postlingual 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) CI 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 CI 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 Rasch analysis (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 composed of 29 items that have been selected to minimise respondent burden while 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 CI 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 a minimum sample size of 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](#).

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 CI centre. Each questionnaire pack will be coded with a unique identifier and 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—classical test theory (CTT) and Rasch measurement theory.^{29 35} Traditional psychometric analyses (eg, 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.³⁶ However, modern measurement techniques such as RA are increasingly being reported alongside traditional analyses in studies of PROM development and validation (eg, refs [37 38](#)).

CTT is based on the assumption that every observed score is a function of an individual's true score and random error.³⁹ The assumptions underpinning CTT differ from those underpinning the Rasch model. For example, it has been argued that CTT cannot be tested adequately as it is based on definitions rather than assumptions which can be proven true or false. This is in contrast to modern measurement theory (ie, RA) which generates

Table 1 Study eligibility criteria for recruitment of participants

Inclusion criteria	Exclusion criteria
<ul style="list-style-type: none"> ▶ Adults (persons ≥18 years of age). ▶ Postlingual severe-profound SNHL ▶ Candidate for cochlear implantation according to UK criteria specified by the National Institute for Health and Care Excellence (2009) or cochlear implant recipient.⁵⁷ ▶ Proficient readers/writers of English. ▶ Capacity to give informed consent. ▶ No additional medical conditions precluding the participant's ability to self-complete the questionnaires 	<ul style="list-style-type: none"> ▶ Prelingual severe-to-profound SNHL (ie, 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 (eg, British Sign Language).

SNHL, sensorineural hearing loss.

assumptions that can be proven true or false.⁴⁰ 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.⁴¹ 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 (ie, patients) according to the amount of the latent target construct (ie, listening effort) they possess and for ordering items that measure the target construct according to their difficulty.³⁵ This method allows non-linear (ie, ordinal) raw data to be converted to a linear (ie, interval) scale, which can then be evaluated through the use of parametric statistical tests.⁴² By contrast, CTT methods yield measures that produce ordinal rather than interval level data. This principle has implications for the interpretation of test scores since difference scores and change scores are most meaningful when interval level of measurement is used.^{35 40}

A further limitation of CTT is that the performance of a test is dependent on the sample in which that test is assessed.⁴⁰ This renders its psychometric properties (ie, 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.⁴³

Despite these limitations, CTT methods continue to be widely used in studies of instrument validation and are included in the COSMIN standards.^{28 40} Indeed, some properties (eg, acceptability, scaling assumptions) can only be evaluated using CTT methods.³⁶ 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 (V.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 (V.4.1.0) software. Psychometric analyses applying CTT will be conducted using SPSS V.22.0 licensed to Swansea University.

Phase 1: item and 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.^{35 44} 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.⁴⁵

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.³⁵ Mean square standardised 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 (ie, 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.⁴⁶

Assessing differential item functioning

Differential item functioning (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 (ie, within error) across all appropriate measurement conditions.³⁵ 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.⁴⁴ The methodology proposed by Zumbo using logistic regression DIF tests for significance (ie, X^2 two df test) and magnitude of DIF by computing the R-squared effect for both uniform and non-uniform DIF will be applied.⁴⁷ 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.^{35 48}

Assessing response scale ordering

The response options of an instrument (ie, number of categories and their definitions) are critical to its reliability and validity.⁴⁹ 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.³⁵ 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.⁵⁰ These investigations will show whether the response options selected for the LEQ-CI are sufficient or should be adjusted 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'.⁴⁴ The means and SD of items and persons should match closely.³⁵

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 variable³⁵ and is considered to be a measure of internal consistency reliability.⁵¹ 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.⁵²

Phase 1: psychometric evaluation using CTT

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 (ie, 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 SD. 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).⁵³ Targeting will be assessed following item refinement using RA. CTT will be used to 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 as 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 ≥ 0.70 and ≤ 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 (ie, listening effort). These state the relationship between the instrument and other measures, as well as the expected differences between the scores attained by different subgroups 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 (ie, 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,⁵⁴ a measure of fatigue used in other studies investigating LE and fatigue in individuals with HL.⁵⁵ As LE and fatigue are similar but unrelated constructs,¹⁶ we anticipate a moderate positive correlation between 0.30 and 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,⁵⁶ a measure of quality of life in CI patients. A small positive correlation of between 0.30 and 0.50 is

anticipated as these measures may be considered to assess similar, but unrelated constructs.²³

ETHICS AND DISSEMINATION

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

Author affiliations

¹Swansea University Medical School, Swansea University, Swansea, UK

²South Wales Cochlear Implant Programme, Bridgend, UK

³Australian Institute of Health Innovation, Macquarie University, Sydney, New South Wales, Australia

⁴Department of Linguistics (Audiology Section), Macquarie University, Sydney, New South Wales, Australia

⁵The HEARING CRC, Melbourne, Australia

Contributors SEH, HAH, AW and FR conceived the study. SEH wrote the protocol and this manuscript. HAH, AW, FR, IB and 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 SEH.

Competing interests None declared.

Patient consent for publication Not required.

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

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

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 KJ, 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;53:433–45.
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 EL, 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. *Health Rep* 2016;27:3–12.
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;62:1841–3.
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 edn. 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-making. 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 BW. 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;1.
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;39:470–4.
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.
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;7: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 SA. Development of the communication 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, 2018:1–78.
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–27.
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 systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1171–9.
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 listening 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. Bond TG, Fox CM. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. 3rd ed. London: Routledge, 2015.
36. 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.
37. 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.
38. 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* 2019;40:328–39.

39. Tractenberg RE. Classical and modern measurement theories, patient reports, and clinical outcomes. *Contemp Clin Trials* 2010;31:1–3.
40. Cano SJ, Hobart JC. The problem with health measurement. *Patient Prefer Adherence* 2011;5:279–90.
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. Boone WJ. Rasch analysis for instrument development: why, when, and how? *CBE Life Sci Educ* 2016;15:rm4.
43. 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;19:1–188.
44. Tadić 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.
45. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol Assess* 1995;7:286–99.
46. Wright L. Reasonable mean-square fit values. *Rasch Meas Trans* 1994;8.
47. 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. 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](http://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).
48. 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.
49. 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.
50. Linacre JM. Categories Disordering (disordered categories) vs. Threshold Disordering (disordered thresholds). *Rasch Meas Trans* 1999;13:675.
51. 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.
52. Tennant A, Conaghan PG. The Rasch measurement 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.
53. 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.
54. Michielsen HJ, De Vries J, Van Heck GL, et al. Examination of the dimensionality of fatigue. *Eur J Psychol Assess* 2004;20:39–48.
55. Alhanbali S, Dawes P, Lloyd S, et al. Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear Hear* 2017;38:e39–e48.
56. 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.
57. 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).