

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 1
4
5 2
6
7
8 3
9
10 4 Study protocol for the validation of a new patient-reported outcome measure (PROM) of
11
12 5 listening effort in cochlear implantation: The Listening Effort Questionnaire-Cochlear
13
14 6 Implant (LEQ-CI)
15
16
17 7

18
19 8 Sarah E. Hughes^{1,2}

20
21 9 Frances Rapport^{3,6}

22
23
24 10 Alan Watkins²

25
26 11 Isabelle Boisvert^{3,4,5}

27
28 12 Catherine M. McMahon^{3,4,5}

29
30 13 Hayley A. Hutchings²
31
32
33 14
34
35 15
36
37

38 16 ¹South Wales Cochlear Implant Programme, ABMU Health Board, Bridgend, Wales, United
39
40 17 Kingdom

41
42 18 ²Patient and Population Health and Informatics, Swansea University Medical School, Swansea,
43
44 19 Wales, United Kingdom

45
46 20 ³Centre for Implementation of Hearing Research, Macquarie University, Sydney, Australia

47
48 21 ⁴Department of Linguistics (Audiology Section), Macquarie University, Sydney, Australia

49
50 22 ⁵The HEARing Cooperative Research Centre, Australia

51
52 23 ⁶Australian Institute of Health Innovation, Macquarie University, Sydney, Australia
53
54 24
55
56 25
57
58
59
60

1
2
3 26
4
5
6 27 Corresponding Author:
7
8 28 Sarah E. Hughes
9
10 29 South Wales Cochlear Implant Programme
11
12 30 Princess of Wales Hospital
13
14 31 Coity Road, Bridgend
15
16 32 Wales, United Kingdom CF31 5EU
17
18
19 33 E-mail: sarah.hughes@wales.nhs.uk
20
21 34 Phone: +44 1656 752192
22
23 35 Fax: +44 1656 752192
24
25
26 36
27
28 37
29
30 38
31
32
33 39 Word count (excluding title page, abstract, references): 3239
34
35 40 Key words: hearing loss, cochlear implant, listening effort, patient-reported outcome measure,
36
37 41 validation, audiology
38
39
40 42
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 43 **ABSTRACT**

4
5 44 **Introduction:** Hearing loss (HL) affects 11 million people in the UK. Associated with HL is a
6
7
8 45 requirement for high levels of effort when listening. Listening effort (LE) may be defined as
9
10 46 the attentional and cognitive resources needed to listen. A number of hearing loss-specific
11
12 47 patient-reported outcome measures (PROMs) are used routinely in the audiology/cochlear
13
14 48 implant (CI) clinic; however, none adequately address the requirement for LE and the impact
15
16
17 49 sustained effortful listening has on individuals' quality of life. This UK-based study aims to
18
19 50 undertake an initial psychometric validation of a newly-developed, disease-specific PROM of
20
21 51 LE, the Listening Effort Questionnaire-Cochlear Implant (LEQ-CI), in the UK population of
22
23 52 adult cochlear implant candidates and recipients. The study will also establish whether further
24
25 53 refinements are required to improve its measurement properties.

26
27
28 54 **Methods and analysis:** This UK-based study is a multi-phase validation study that has been
29
30 55 designed in accordance with the internationally recognised COSMIN standards. In Phase 1,
31
32 56 adult CI candidates and recipients (n = 250) will self-complete a paper-and-pencil version of
33
34 57 the draft LEQ-CI. Participants' responses to the LEQ-CI's items will be assessed to establish
35
36 58 unidimensionality and Rasch Analysis will evaluate item and scale functioning. Classical Test
37
38 59 Theory (CTT) will assess acceptability/data completeness, scaling assumptions, targeting, and
39
40 60 internal consistency reliability. Phase 1 results will inform adjustments to the items, scale(s),
41
42 61 and response options to produce a refined version of the LEQ-CI. In Phase 2, a new sample (n
43
44 62 = 100) will self- complete the refined LEQ-CI and four comparator PROMs to assess construct
45
46
47 63 validity.

48
49
50 64 **Ethics and dissemination:** This study has been approved by the Abertawe Bro Morgannwg
51
52 65 University (ABMU) Health Board/Swansea University Joint Study Review Committee (JSRC)
53
54 66 and the Newcastle and North Tyneside 2 Research Ethics Committee (REC), Ref: 18/NE/0320.
55
56
57
58
59
60

1
2
3 67 Findings will be disseminated in high-quality peer-reviewed journals, conference
4
5 68 presentations, and SEH's doctoral dissertation. (299 words)
6
7

8 **ARTICLE SUMMARY**

9 **Strengths and limitations**

- 10
11
12 71 • The LEQ-CI is the first PROM developed specifically to assess perceived listening effort
13
14 72 in cochlear implant candidates and recipients.
15
16
17 73 • The proposed study conforms to international consensus standards on best practice of
18
19 74 studies of instrument development and validation, the COsensus-based Standards for the
20
21 75 selection of health-status Measurement INstruments (COSMIN).
22
23
24 76 • The use of CTT and Rasch Analysis will enable a robust initial assessment of the LEQ-
25
26 77 CI's measurement characteristics at both item and scale level.
27
28
29 78 • The conceptual framework underpinning the LEQ-CI is based on an explanatory model
30
31 79 developed from current theoretical frameworks and the patient perspective. Assessment of
32
33 80 the LEQ-CI's measurement properties will provide early evidence of the validity of the
34
35 81 proposed model.
36
37
38 82 • Instrument validation is an iterative process to build a body of evidence relating to the
39
40 83 quality of an instrument's measurement properties. Further studies that assess the
41
42 84 measurement characteristics of LEQ-CI will be required.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

87 INTRODUCTION

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

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

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

1
2
3 112 Behavioural performance measures (e.g., the dual-task paradigm) as measures of listening
4
5 113 effort are also well-reported in the literature.[17] However, the complex nature of the listening
6
7 114 effort construct means these tools may not be capturing all aspects of the construct and may
8
9 115 not be appropriate for assessing those properties of listening effort that are important to
10
11 116 individuals when listening in the real world. A PROM has the potential to offer a clinically
12
13 117 viable and alternative view on listening effort from the patient perspective. A systematic review
14
15 118 of PROMs considered to measure perceived listening effort established that a number of self-
16
17 119 report questionnaires have been developed that include individual items considered to assess
18
19 120 listening effort but, notably, found no validated self-report instruments developed specifically
20
21 121 to measure perceived listening effort in a manner consistent with current theoretical
22
23 122 frameworks.[12,18]

123 AIM

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

133 OBJECTIVES

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

- 1
2
3 137 • To undertake an initial assessment of the LEQ-CI's psychometric properties, applying CTT
4
5 138 to assess acceptability, scaling assumptions, targeting, and reliability.
6
7
8 139 • To assess the construct validity of the LEQ-CI in the population of adults with severe-
9
10 140 profound, post-lingual HL who are CI candidates or recipients.
11
12
13 141

142 **METHODS AND ANALYSIS**

143 **Study Setting and Patient Involvement**

144 This study is a UK-based multi-phase study to validate the measurement properties of
145 a new PROM, the LEQ-CI. The planned study will take place over a 12-month period and has
146 been co-produced by the study team with input from two lay members, both CI recipients, from
147 the study's Research Management Group. Lay members reviewed and provided feedback on
148 study design, participant documents and iterations of the LEQ-CI.
149

150 **Development of the LEQ-CI**

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

1
2
3 161 categories ensure a broad range of response options. Item responses are summed to produce a
4
5 162 simple total score.
6
7
8 163

10 164 **Sample size**

12 165 The study sample will be representative of the population of adults with acquired, post-
13
14 166 lingual severe-to-profound sensorineural hearing loss (SNHL) referred for cochlear
15
16
17 167 implantation in the UK. A total study sample of 350 participants will be recruited. In Phase 1,
18
19 168 a cohort of 250 participants will be recruited from multiple National Health Service (NHS)
20
21 169 cochlear implant centres. There are no general criteria for determination of sample sizes in
22
23 170 studies of PROM validation and sample sizes are, in part, dependent on the psychometric
24
25 171 characteristics being assessed.[22,23] Mokkink et al. recommend greater than 200 respondents
26
27 172 when undertaking RA and seven times the number of items for purposes of undertaking
28
29 173 assessment of unidimensionality.[22] Linacre recommends a sample size of 250 respondents
30
31 174 for definitive item calibration using RA.[24] The LEQ-CI is comprised of 27 items that have
32
33 175 been selected to minimise respondent burden whilst allowing for adequate sampling of relevant
34
35 176 constructs associated with listening effort. Therefore, a minimum sample of 250 participants
36
37 177 is considered sufficient for undertaking both assessment of unidimensionality and RA of the
38
39 178 LEQ-CI. In Phase 2, a new cohort of 100 participants fulfilling the same eligibility criteria as
40
41 179 Phase 1 will be recruited. Hobart et al. recommend greater than 80 participants for assessment
42
43 180 of construct validity.[25]
44
45
46
47
48
49
50

51 182 **Recruitment and data collection**

53 183 The participant eligibility criteria are the same for both phases of the study and are
54
55 184 presented in Table 1.
56
57

58 185 Table 1. Study eligibility criteria for recruitment of participants
59
60

Inclusion Criteria	Exclusion Criteria
<ul style="list-style-type: none"> • Adults (persons \geq 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 	<ul style="list-style-type: none"> • 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.

186

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

1
2
3 192 participant anonymity, eligibility screening and the study documents will be mailed to
4
5 193 prospective participants by a member of the clinical team at each participating cochlear implant
6
7 194 centre. To maintain participant anonymity, each questionnaire pack will be coded with a unique
8
9 195 identifier. No personal identifiable information will be retained by the study team.
10
11
12
13

196

14 197 **Statistical Analysis**

16
17 198 Study data will be managed using the online clinical data management programme,
18
19 199 REDCap (Version 7.2.1, Vanderbilt University), licensed to the Swansea Trials Unit, Swansea
20
21 200 University. RA will be used to assess item and scale structure of the LEQ-CI using Winsteps
22
23 201 (Version 4.1.0) software. Psychometric analyses applying CTT will be conducted using the
24
25 202 Statistical Package for Social Sciences (SPSS), Version 22.0 licensed to Swansea University.
26
27
28
29

203

30 204 Phase 1: Item & Scale Refinement using Rasch Measurement Theory

31
32
33 205 As a form of modern measurement theory, RA is being applied increasingly to the
34
35 206 development of PROMs as a complement to CTT, the traditional method of psychometric
36
37 207 evaluation. Used to improve the precision of assessment instruments.[27] RA allows
38
39 208 instrument designers to empirically assess the behaviour of both items and response categories.
40
41 209 Central to the Rasch model is a method for ordering persons (i.e., patients) according to the
42
43 210 amount of the latent target construct (i.e., listening effort) they possess and for ordering items
44
45 211 that measure the target construct according to their difficulty.[28] This method allows non-
46
47 212 linear (ordinal) raw data to be converted to a linear scale (interval), which can then be evaluated
48
49 213 through the use of parametric statistical tests.[27] The Rasch model considers how well the
50
51 214 observed data fits the measurement model, unlike CTT which considers how well the model
52
53 215 describes the data.[29] Because the Rasch model is based on theory and is independent of any
54
55 216 data set, any discrepancies between the scale data and the Rasch model requirements are
56
57
58
59
60

1
2
3 217 indicative of anomalies in the scale as a measurement instrument. These discrepancies provide
4
5 218 diagnostic information that serves as a basis for understanding and empirical improvement of
6
7
8 219 the instrument at both item and scale-level.[30]
9

220

221 *Assessing unidimensionality*

14 222 The Rasch measurement model assumes unidimensionality, defined as the
15
16
17 223 measurement of a single latent construct.[28,31] Therefore, prior to undertaking RA, factor
18
19 224 analysis will be undertaken to assess the underlying structure of the LEQ-CI and establish the
20
21 225 unidimensionality of its (sub)scales.[32]
22

226

227 *Assessing item fit*

28 228 In RA, item fit refers to the degree of mismatch between the pattern of actual observed
29
30 229 responses and the Rasch modelled expectations. Specifically, it refers to the pattern for each
31
32
33 230 item across persons investigated by examining item infit and outfit statistics.[28] Mean square
34
35 231 standardized residuals (MNSQ) will be used to assess fit with MNSQ residuals within the 0.5–
36
37 232 1.5 range considered acceptable for productive measurement. Mean square values less than 0.5
38
39 233 indicate overfit (i.e., the items are too predictable relative to the Rasch model), while mean
40
41 234 square values greater than 1.5 are indicative of too much noise (randomness) relative to the
42
43
44 235 Rasch model.[33]
45

236

237 *Assessing differential item functioning (DIF)*

51 238 DIF is an indication of the loss of invariance across subsamples of respondents. The
52
53 239 presence of DIF will be an indicator of potential problems with an item since item and person
54
55 240 measures on a unidimensional instrument should remain invariant (i.e., within error) across all
56
57 241 appropriate measurement conditions.[28] DIF will be examined for key demographic variables
58
59
60

1
2
3 242 such as age and sex. The standard threshold of > 1 logit will be used as an indicator of DIF.[31]
4
5 243 The methodology proposed by Zumbo using logistic regression DIF tests for significance (i.e.,
6
7 244 Chi-squared two degrees of freedom test) and magnitude of DIF by computing the R-squared
8
9 245 effect for both uniform and non-uniform DIF will be applied.[34] If items are found to have
10
11 246 DIF they will either be considered candidates for removal or examined for adjustment of DIF
12
13 247 and re-evaluated, thus reflecting the iterative nature of instrument validation.[28,35]
14
15
16
17 248

19 249 *Assessing response scale ordering*

21 250 The response options of an instrument (i.e., number of categories and their definitions)
22
23 251 are critical to its reliability and validity.[36] The Rasch model will enable us to show
24
25 252 empirically how respondents use the LEQ-CI's rating scale informing future iterations of the
26
27 253 LEQ-CI to ensure it yields high quality data.[28] Response category ordering will be assessed
28
29 254 using Rasch probability curves and there will be an examination of the data for category
30
31 255 disordering and threshold disordering.[37] These investigations will show whether the
32
33 256 response options selected for the LEQ-CI are sufficient or should be collapsed to provide better
34
35 257 coverage of the latent trait.
36
37
38
39
40 258

42 259 *Assessing the targeting of persons and items*

44 260 Targeting explores whether the instrument has a distribution of items that matches the
45
46 261 range of the respondents' latent trait. This will be done by examining the item-person threshold
47
48 262 distribution map, which illustrates a relative position of "item difficulty" to "person
49
50 263 ability".[31] The means and standard deviations of items and persons should match closely.[28]
51
52
53
54 264

56 265 *Assessing reliability*

1
2
3 266 The reliability of the LEQ-CI will be examined by observing the person separation
4
5 267 index (PSI). The PSI is an estimate of the spread or separation of persons on the measured
6
7
8 268 variable [28] and is considered to be a measure of internal consistency reliability.[38] It is a
9
10 269 measure of the scale's ability to separate the study sample. A $PSI > 0.7$ will be considered an
11
12 270 adequate measure of reliability.[39]

13
14
15 271

16 17 272 Phase 1: Psychometric evaluation using Classical Test Theory

18
19 273 Initial assessment of the LEQ-CI's psychometric properties will be undertaken
20
21 274 applying principles of CTT that comply with the current "gold standard" proposed by the
22
23
24 275 COSMIN group.[22] As a theory of measurement, CTT seeks to evaluate the reliability and
25
26 276 validity of a scale and has been the dominant approach used in the development and validation
27
28 277 of outcome measures. CTT is based on the assumption that every observed score is a function
29
30 278 of an individual's true score and random error.[40] The assumptions underpinning CTT differ
31
32
33 279 from those underpinning the Rasch model. Notably, CTT focuses only on the total test score
34
35 280 (i.e., the summary of item scores) to establish reliability and validity. Item-level measurement
36
37 281 properties are not considered; therefore, limiting the use of CTT as a method to assess the
38
39 282 performance of individual items.[41] For this study, item function will be assessed first using
40
41
42 283 RA followed by complementary psychometric analyses using CTT. Specifically, CTT will be
43
44 284 used to evaluate the LEQ-CI for its acceptability, targeting, scaling assumptions, and internal
45
46 285 consistency reliability.

47
48
49 286

50 51 287 *Assessing acceptability and data completeness*

52
53 288 Acceptability and data completeness will establish extent to which scale items are
54
55 289 scored and total scores can be computed. Assessment of the completeness of item and scale-
56
57
58 290 level data (i.e., missing or incomplete data for items and sample) including frequency of
59
60

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

293

294 *Assessing scaling assumptions*

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

298

299 *Assessing targeting*

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

305

306 *Assessing internal consistency reliability*

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

310 • Inter-item correlation – Calculating inter-item correlations will provide an indication
311 whether an item is part of a (sub)scale. Correlations should fall between 0.2 and 0.5. Items
312 which have a correlation greater than 0.7 may be considered to measure the same thing,
313 making one item a candidate for deletion.[23]

314 • Item-total correlation – Calculating item-total correlations will assess whether the LEQ-
315 CI’s items discriminate patients on the listening effort construct. Items that show an item-

1
2
3 316 total correlation of less than 0.3 will be considered as contributing little to the LEQ-CI in
4
5 317 terms of discriminating between individuals with high versus low levels of listening
6
7 318 effort.[23] These items will be considered candidates for deletion.

9
10 319 • Cronbach's alpha - Internal consistency will be calculated for each subscale of the LEQ-CI
11
12 320 by calculating Cronbach's alpha.[19,45–47] Alpha values ≥ 0.70 and ≤ 0.95 will be
13
14 321 considered good evidence of internal consistency.[19]
15
16

17 322

18
19 323 Phase 2: Establishing construct validity
20

21 324 Construct validity may be defined as the extent to which the scores of an instrument
22
23 325 are a valid measure of the latent construct.[23] Construct validity of the refined LEQ-CI will
24
25 326 be assessed by applying criteria specified by the COSMIN group. Construct validity may be
26
27 327 assessed by testing a priori hypotheses about the relationship between the instrument and
28
29 328 other measures, as well as the expected differences between the scores attained by different
30
31 329 sub-groups of the target population based on the assumption that the LEQ-CI validly
32
33 330 measures the target construct (i.e., listening effort). To establish the construct validity of an
34
35 331 instrument, Mokkink et al. recommend at least 75% of the stated hypotheses are endorsed.
36
37 332 [22]
38
39
40
41
42
43
44

45 334 *Assessing convergent validity*

46
47 335 Concurrent construct validity will be assessed by examining the correlation between
48
49 336 scores on the LEQ-CI with the summed score on the three items considered to measure listening
50
51 337 effort on the Speech, Spatial, and Qualities of Hearing questionnaire (SSQ).[48] As no
52
53 338 validated measure of listening effort has been identified as a suitable comparator PROM these
54
55 339 items were selected to assess construct validity as the SSQ has good evidence of being a well-
56
57 340 validated instrument across multiple studies.[18] We hypothesise that a strong positive
58
59
60

1
2
3 341 correlation > 0.50 will be observed between measures. We further predict a moderate positive
4
5 342 correlation (0.30 – 0.50) between the LEQ-CI and SSQ total score as LE may be considered to
6
7 343 be a component of hearing disability, the construct measured by the SSQ.[5]
8
9

10 344

11 12 345 *Assessing discriminant validity*

13
14 346 Discriminant (i.e., divergent) validity is an assessment of a measure's ability to
15
16 347 discriminate between dissimilar constructs.[19] It will be assessed by the examining the
17
18 348 correlation between scores on the LEQ-CI and the Fatigue Assessment Scale (FAS),[49] a
19
20 349 measure of fatigue used in other studies investigating LE and fatigue in individuals with
21
22 350 HL.[50] As LE and fatigue are similar but unrelated constructs we anticipate a moderate
23
24 351 positive correlation between 0.30 – 0.50.
25
26
27

28 352 Further assessment of discriminant validity will be undertaken by examining the
29
30 353 correlation between scores on the LEQ-CI and the Nijmegen Cochlear Implant Questionnaire
31
32 354 (NCIQ),[51] a measure of quality of life in CI patients. A small positive correlation of between
33
34 355 0.30 – 0.50 is anticipated as these measures may be considered to assess similar, but unrelated
35
36 356 constructs. [22]
37
38
39

40 357

41 42 358 **ETHICS AND DISEMINATION**

43
44 359 This study has received ethical approval from the NHS Research Ethics Committee
45
46 360 (REC): Newcastle and North Tyneside 2 (Ref: 18/NE/0320). Study findings will be
47
48 361 disseminated through publication in peer-reviewed journals, conference presentations, and in
49
50 362 the lead author's (SEH) doctoral dissertation.
51
52
53

54 363

1
2
3 364 **Author contributions:** SEH, HAH, AW and FR conceived the study. SH wrote the protocol
4
5 365 and this manuscript. HAH, AW, FR, IB, CMM provided critical review of the protocol and this
6
7 366 manuscript. All authors read and approved the final manuscript.

8
9
10 367 **Funding:** This project is supported by an Abertawe Bro Morgannwg University Health Board
11
12 368 Pathway to Portfolio grant (R&D Pathway June002) awarded to Mrs. Sarah Hughes.

13
14 369 **Competing interests:** The authors have no competing interests to declare.

15
16
17 370 **Provenance and peer review:** Not commissioned; externally peer reviewed.

18
19 371 **Open Access:** This is an Open Access article distributed in accordance with the terms of the
20
21 372 Creative Commons Attribution (CC BY NC 4.0) license, which permits others to distribute,
22
23 373 remix, adapt and build upon this work, for non-commercial use, provided the original article is
24
25 374 properly cited. To view a copy of this license, visit [http://creativecommons.org/licenses/by-](http://creativecommons.org/licenses/by-nc/4.0/)
26
27
28 375 [nc/4.0/](http://creativecommons.org/licenses/by-nc/4.0/).

29
30
31 376

32
33 377 **REFERENCES**

- 34
35 378 1 Hutchings HA, Alrubaiy L. Patient-reported outcome measures in routine clinical
36
37 379 care : The PROMise of a better future? *Dig Dis Sci* Published Online First: 2017.
38
39 380 doi:10.1007/s10620-017-4658-z
- 40
41
42 381 2 Ousey K, Cook L. Understanding patient reported outcome measures (PROMs). *Br J*
43
44 382 *Community Nurs* 2011;16:80–2.
- 45
46
47 383 3 Meadows KA. Patient-reported outcome measures: An overview. *Br J Community*
48
49 384 *Nurs* 2011;16(3):146–51.
- 50
51
52 385 4 Devlin NJ, Appleby J, Buxton M, et al. Getting the Most Out of PROMs: Putting
53
54 386 Health Outcomes at the Heart of NHS Decision-Making. London: The King's Fund
55
56 387 2010.
- 57
58 388 5 Noble W. Self-Assessment of Hearing. 2nd ed. San Diego: Plural Publishing 2013.
59
60

- 1
2
3 389 6 McGarrigle R, Munro K, Dawes P, et al. Listening effort and fatigue: What exactly are
4 we measuring? A British Society of Audiology Cognition in Hearing Special Interest
5 390 Group ‘white paper’. *Int J Audiol* 2014;53:433-40.
6
7 391
8
9
10 392 7 Mackersie CL, MacPhee IX, Heldt EW. Effects of hearing loss on heart rate variability
11 and skin conductance measured during sentence recognition in noise. *Ear Hear*
12 393 2015;36:145-54.
13
14 394
15
16 395 8 Zekveld AA, George ELJ, Houtgast T, et al. Cognitive abilities relate to self-reported
17 hearing disability. *J Speech Lang Hear Res* 2013;56:1364–72.
18
19 396
20
21 397 9 Ramage-Morin PL. Hearing difficulties and feelings of social isolation among
22 Canadians aged 45 or older. *Health Rep* 2016;27:3–12.
23
24 398
25
26 399 10 Nachtegaal J, Kuik DJ, Anema JR, et al. Hearing status, need for recovery after work,
27 and psychosocial work characteristics: results from an internet-based national survey
28 400 on hearing. *Int J Audiol* 2009;48:684–91.
29
30 401
31
32 402 11 Miles K, McMahon C, Boisvert I, et al. Objective assessment of listening effort:
33 Coregistration of pupillometry and EEG. *Trends Hear* 2017;21:233121651770639.
34 403 doi:10.1177/2331216517706396
35
36 404
37
38 405 12 Pichora-Fuller MK, Kramer SE, Eckert MA, et al. Hearing impairment and cognitive
39 energy: The Framework for Understanding Effortful Listening (FUEL). *Ear Hear*
40 406 2016;37(Suppl1):5S–27S.
41
42 407
43
44 408 13 Zekveld AA, Heslenfeld DJ, Johnsrude IS, et al. The eye as a window to the listening
45 brain: Neural correlates of pupil size as a measure of cognitive listening load.
46 409 *Neuroimage* 2014;101:76–86.
47
48 410
49
50 411 14 Winn MB, Edwards JR, Litovsky RY. The impact of auditory spectral resolution on
51 listening effort revealed by pupil dilation. *Ear Hear* 2015;36:e153–65.
52 412 doi:10.1097/AUD.0000000000000145
53
54
55
56
57
58
59
60

- 1
2
3 414 15 Bernarding C, Strauss DJ, Hannemann R, et al. Quantification of listening effort
4
5 415 correlates in the oscillatory EEG activity: A feasibility study. *Conf Proc IEEE Eng*
6
7 416 *Med Biol Soc* 2012;2012:4615–8.
- 8
9
10 417 16 McMahon CM, Boisvert I, de Lissa P, et al. Monitoring alpha oscillations and pupil
11
12 418 dilation across a performance-intensity function. *Front Psychol* 2016;7:745.
13
14 419 doi:10.3389/fpsyg.2016.00745
- 15
16
17 420 17 Gagné J, Besser J, Lemke U. Behavioral assessment of listening effort using a dual-
18
19 421 task paradigm. *Trends Hear* 2017;21:1–25. doi:10.1177/2331216516687287
- 20
21 422 18 Hughes SE, Hutchings HA, Peters S, et al. A systematic review and narrative synthesis
22
23 423 of the measurement properties of patient-reported outcome measures (PROMs) used to
24
25 424 assess listening effort in hearing loss. *British Society of Audiology Conference 2017*.
26
27 425 Harrogate: 2017.
- 28
29
30 426 19 Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their
31
32 427 Development and Use. 4th ed. Oxford University Press 2008.
- 33
34
35 428 20 Hughes SE, Hutchings HA, Rapport FL, et al. Social connectedness and perceived
36
37 429 listening effort in adult cochlear implant users: A Grounded Theory to establish
38
39 430 content validity for a new patient-reported outcome measure. *Ear Hear* 2018;39:922–
40
41 431 34.
- 42
43
44 432 21 Hughes SE, Rapport FL, Boisvert I, et al. Patient-reported outcome measures
45
46 433 (PROMs) for assessing perceived listening effort in hearing loss: Protocol for a
47
48 434 systematic review. *BMJ Open* 2017;7:e014995. doi:10.1136/bmjopen-2016-014995
- 49
50
51 435 22 Mookink LB, de Vet HCW, Prinsen DL, et al. COSMIN risk of bias checklist for
52
53 436 systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;
54
55 437 27:1171-1179.
- 56
57
58 438 23 de Vet H, Terwee CB, Mookink LB, et al. Measurement in Medicine: A Practical
59
60

- 1
2
3 439 Guide. Cambridge:Cambridge University Press 2011.
4
5 440 24 Linacre JM. Sample size and item calibration or person measure stability. *Rasch Meas*
6
7 441 *Trans* 1994;7:328.
8
9 442 25 Hobart JC, Cano SJ, Warner TT, *et al*. What sample sizes for reliability and validity
10
11 443 studies in neurology? *J Neurol* 2012;259:2681–94.
12
13 444 26 National Institute for Health and Clinical Excellence (NICE). NICE technology
14
15 445 appraisal guidance 166: Cochlear implants for children and adults with severe to
16
17 446 profound deafness. 2009. www.guidance.nice.org.uk/TA166/pdf/English (accessed 18
18
19 447 May 2018).
20
21
22 448 27 Boone WJ. Rasch analysis for instrument development: Why, when, and how? *CBE*
23
24 449 *Life Sci Educ* 2016;15:pii-rm4.
25
26 450 28 Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the
27
28 451 Human Sciences 3rd Ed. New York: Routledge 2015.
29
30 452 29 Petrillo J, Cano SJ, McLeod LD, *et al*. Using Classical Test Theory, Item Response
31
32 453 Theory, and Rasch Measurement Theory to evaluate patient-reported outcome
33
34 454 measures: A comparison of worked examples. *Value Health* 2015;18:25–34.
35
36 455 30 Ball S, Vickery J, Hobart J, *et al*. The Cannabinoid Use in Progressive Inflammatory
37
38 456 brain Disease (CUPID) trial: A randomised double-blind placebo-controlled parallel-
39
40 457 group multicentre trial and economic evaluation of cannabinoids to slow progression
41
42 458 in multiple sclerosis. *Health Technol Assess* 2015;19:1–187. doi:10.3310/hta19120
43
44 459 31 Tadic V, Cooper A, Cumberland P, *et al*. Measuring the quality of life of visually
45
46 460 impaired children: First stage psychometric evaluation of the novel VQoL_CYP
47
48 461 instrument. *PLoS One* 2016;11:e0146225.
49
50 462 32 Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical
51
52 463 assessment instruments. *Psychol Assess* 1995;7:286–99.
53
54
55
56
57
58
59
60

- 1
2
3 464 33 Wright B, Linacre JM. Reasonable mean-square fit values. *Rasch Anal Trans* 1994;8:
4
5 465
6 370.
7
8 466 34 Zumbo B. A handbook on the theory and methods of differential item functioning
9
10 467 (DIF). *Ottawa Natl Def Hqrs* 1999:1–57.
11
12 468 35 Teresi JA, Ramirez M, Jones RN, et al. Modifying measures based on differential item
13
14 469 functioning (DIF) impact analysis. *J Aging Health* 2012;24:1044–76.
15
16
17 470 36 Johnston M, Dixon D, Hart J, et al. Discriminant content validity: A quantitative
18
19 471 methodology for assessing content of theory-based measures, with illustrative
20
21 472 applications. *Br J Health Psychol* 2014;19:240–57.
22
23
24 473 37 Linacre JM. Category disordering (disordered categories) vs. threshold disordering
25
26 474 (disordered thresholds). *Rasch Meas Trans* 1999;13:675.
27
28
29 475 38 Gorecki CA. *The development and validation of a patient-reported outcome measure*
30
31 476 *of health-related quality of life for patients with pressure ulcers: PUQOL Project*
32
33 477 University of Leeds 2011.
34
35
36 478 39 Tennant A, Conaghan, PG. The measurement model model in rheumatology: What is
37
38 479 it and why use it? When should it be applied, and what should one look for in a Rasch
39
40 480 paper? *Arthritis Rheum* 2007;57:1358–62.
41
42
43 481 40 Tractenberg RE. Classical and modern measurement theories, patient reports and
44
45 482 clinical outcomes. *Contemp Clin Trials* 2010;31:1–3.
46
47
48 483 41 da Rocha NS, Chachamovich E, de Almeida Fleck MP, et al. An introduction to Rasch
49
50 484 analysis for psychiatric practice and research. *J Psychiatr Res* 2013;47:141–8.
51
52
53 485 42 McHorney CA, Ware, Jr. JE, Lu FR, et al. The MOS 36-Item Short-Form Health
54
55 486 Survey (SF-36): III . Tests of data quality, scaling assumptions, and reliability across
56
57 487 diverse patient groups. *Med Care* 1994;32:40–66.
58
59
60 488 43 Gorecki C, Brown JM, Cano S, et al. Development and validation of a new patient-

- 1
2
3 489 reported outcome measure for patients with pressure ulcers: The PU-QOL instrument.
4
5
6 490 *Health Qual Life Outcomes* 2013;11:1.
7
8 491 44 Mokkink LB, Terwee CB, Patrick DL, et al. COSMIN checklist manual. Amsterdam:
9
10 492 COSMIN 2012.
11
12 493 45 Streiner DL. Starting at the beginning: An introduction to coefficient alpha and
13
14 494 internal consistency. *J Pers Assess* 2003;80:99–103.
15
16
17 495 46 Streiner DL. Being inconsistent about consistency: When coefficient alpha does and
18
19 496 doesn't matter. *J Pers Assess* 2003;80:217–22.
20
21 497 47 Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic
22
23 498 reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147-1157.
24
25
26 499 48 Gatehouse S, Noble W. The Speech, Spatial and Qualities of Hearing scale (SSQ). *Int*
27
28 500 *J Audiol* 2004;43:85–99.
29
30
31 501 49 Michielsen HJ, De Vries J, Van Heck GL, et al. Examination of the dimensionality of
32
33 502 fatigue. *Eur J Psychol Assess* 2004;20:39–48.
34
35
36 503 50 Alhanbali S, Dawes P, Lloyd S, et al. Self-reported listening-related effort and fatigue
37
38 504 in hearing-impaired adults. *Ear Hear* 2017;38:e39–48.
39
40 505 51 Hinderink JB, Krabbe PF, Van Den Broek P. Development and application of a health-
41
42 506 related quality-of-life instrument for adults with cochlear implants: The Nijmegen
43
44 507 cochlear implant questionnaire. *Otolaryngol Head Neck Surg* 2000;123:756-65.
45
46
47 508
48
49
50
51
52
53
54
55
56
57
58
59
60

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts

1
2
3 1
4
5 2
6
7
8 3
9
10 4 Study protocol for the validation of a new patient-reported outcome measure (PROM) of
11
12 5 listening effort in cochlear implantation: the Listening Effort Questionnaire-Cochlear Implant
13
14
15 6 (LEQ-CI)
16
17 7

18
19 8 Sarah E. Hughes^{1,2}

20
21
22 9 Frances Rapport^{3,6}

23
24 10 Alan Watkins²

25
26 11 Isabelle Boisvert^{3,4,5}

27
28 12 Catherine M. McMahon^{3,4,5}

29
30
31 13 Hayley A. Hutchings²
32
33 14
34
35 15
36
37

38 16 ¹South Wales Cochlear Implant Programme, Bridgend, Wales, United Kingdom

39
40 17 ²Patient and Population Health and Informatics, Swansea University Medical School, Swansea,
41
42 18 Wales, United Kingdom

43
44 19 ³Centre for Implementation of Hearing Research, Macquarie University, Sydney, Australia

45
46 20 ⁴Department of Linguistics (Audiology Section), Macquarie University, Sydney, Australia

47
48
49 21 ⁵The HEARing Cooperative Research Centre, Australia

50
51 22 ⁶Australian Institute of Health Innovation, Macquarie University, Sydney, Australia
52
53
54 23
55
56 24
57
58 25
59
60

1
2
3 26 Corresponding Author:
4

5 27 Sarah E. Hughes
6

7 28 South Wales Cochlear Implant Programme
8

9
10 29 Princess of Wales Hospital
11

12 30 Coity Road, Bridgend
13

14 31 Wales, United Kingdom CF31 5EU
15

16
17 32 E-mail: sarah.hughes@wales.nhs.uk
18

19 33 Phone: +44 1656 752192
20

21 34 Fax: +44 1656 752192
22

23
24 35
25

26 36
27

28 37
29

30 38 Word count (excluding title page, abstract, references, figures and tables): 3740
31

32 39 Key words: hearing loss, cochlear implant, listening effort, patient-reported outcome measure,
33

34 40 validation, PROM
35

36
37 41
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

1
2
3 67 Dissemination will be in high-quality journals, conference presentations, and SEH's doctoral
4
5 68 dissertation. (300 words)
6

7 69 **ARTICLE SUMMARY**

10 70 **Strengths and limitations**

- 12 71 • The LEQ-CI is the first PROM developed specifically to assess perceived listening effort
13
14 72 in cochlear implant candidates and recipients.
- 15
16
17 73 • The proposed study conforms to international consensus standards on best practice of
18
19 74 studies of instrument development and validation - the COnsensus-based Standards for the
20
21 75 selection of health-status Measurement INstruments (COSMIN).
- 22
23
24 76 • The use of CTT and RA will enable a robust initial assessment of the LEQ-CI's
25
26 77 measurement characteristics at both item and scale level.
- 27
28
29 78 • The conceptual framework underpinning the LEQ-CI is based on an explanatory model
30
31 79 developed from current theoretical frameworks and the patient perspective. Assessment of
32
33 80 the LEQ-CI's measurement properties will provide early evidence of the validity of the
34
35 81 proposed model.
- 36
37
38 82 • Instrument validation is an iterative process to build a body of evidence relating to the
39
40 83 quality of an instrument's measurement properties. Further studies that assess the
41
42 84 measurement characteristics of LEQ-CI will be required.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 112 patient themselves. PROMs offer a complementary method to current behavioural (e.g., dual
4
5 113 task paradigms) and physiological measures (e.g., pupilometry, fMRI,
6
7 114 electroencephalography) of listening effort. There is a growing body of research to suggest that
8
9 115 listening effort is a multidimensional construct and that these different measures may evaluate
10
11 116 different aspects of this phenomenon.[4,13–16] Using factor analysis, Alhanbali et al. have
12
13 117 shown that hearing level, SNR, dual-task paradigms, pupilometry and EEG (i.e., alpha power
14
15 118 during speech recognition and retention) and self-reported effort tap into different underlying
16
17 119 dimensions of listening effort.[14] Reflecting on this work, it may be argued that PROMs, as
18
19 120 a measure of self-reported effort, have the potential to assess a dimension of listening effort
20
21 121 that is not captured by current behavioural and physiological measures.

22
23
24
25
26 122 Several hearing-specific PROMs have been developed that include items considered to
27
28 123 measure listening effort. A systematic review by the authors identified two PROMs that
29
30 124 measured listening effort and cognitive effort in listening respectively.[17,18] Several PROMs
31
32 125 assessing listening effort at either the item or subscale level (e.g., SSQ, (A)PHAB, CPHI) were
33
34 126 also identified.[19–22] Overall, the review findings found limited evidence of these PROMs'
35
36 127 psychometric measurement properties. The SSQ was identified as the current best candidate
37
38 128 for use as a listening effort PROM based on the extent and quality of its validation when
39
40 129 assessed against the COSMIN criteria.[23] However, one drawback of the SSQ as a measure
41
42 130 of listening effort is a high response burden with only 6% of its items measuring listening
43
44 131 effort. Notably, all of the PROMs identified in this systematic review were developed prior to
45
46 132 publication of the theoretical frameworks and treatises that inform current conceptualisations
47
48 133 of listening effort including the role of motivation on effort expenditure.[2,24–26] Lack of
49
50 134 congruence between these instruments and current frameworks is a limitation of the content
51
52 135 validity of existing PROMs. It is unlikely these instruments capture fully the conceptualisation
53
54 136 of listening effort as presented in these recently published models. As such, there is growing
55
56
57
58
59
60

1
2
3 137 support in the literature for a new PROM that comprehensively measures self-reported listening
4
5 138 effort in hearing loss as it is conceptualised currently.[14,26] To address this situation, the
6
7
8 139 Listening Effort Questionnaire – Cochlear Implant (LEQ-CI) has been developed. The LEQ-
9
10 140 CI is a new hearing-specific PROM measuring perceived listening effort in adults who receive
11
12 141 cochlear implants.
13
14
15 142

17 143 AIMS

19 144 To have confidence that a PROM is providing meaningful information, psychometric
20
21 145 evaluation of its measurement properties must be undertaken to satisfy rigorous criteria.[23,27]
22
23 146 This includes assessment of an instrument's validity (i.e., to what extent does the instrument
24
25 147 measure the construct it purports to measure), its reliability (i.e., the degree to which
26
27 148 measurement is free from error) and its responsiveness (i.e., the ability of an outcome measure
28
29 149 to detect change over time in the construct to be measured).[28] There are several measurement
30
31 150 properties that require assessment and each property needs its own type of study to assess it.
32
33 151 The process of psychometric validation is iterative and represents an accumulation of evidence
34
35 152 over time from multiple studies.[29]
36
37
38
39

40 153 The aim of this study is to conduct the first psychometric validation of the LEQ-CI in
41
42 154 accordance with the internationally recognised COSMIN guidelines.[28,30] Building on
43
44 155 previous work undertaken by the authors to establish the LEQ-CI's content validity,[2,31] the
45
46 156 current study represents a further step towards the provision of a robust self-report measure of
47
48 157 perceived listening effort for use in research and clinical practice.
49
50

51 158 OBJECTIVES

- 52
53
54 159 • To refine the items, response categories, and scale structure of the new LEQ-CI using Rasch
55
56 160 Measurement Theory in an English-speaking sample of adult cochlear implant candidates
57
58 161 and recipients in the UK.
59
60

- 1
2
3 162 • To assess acceptability, scaling assumptions, targeting, and reliability using CTT methods.
4
5
6 163 • To assess the construct validity of the refined LEQ-CI in the population of adults with
7
8 164 severe-profound, post-lingual HL who are CI candidates or recipients.
9

10 165

11
12 166

METHODS AND ANALYSIS

13 167 **Study Setting and Patient Involvement**

14
15
16
17 168 This prospective study is a UK-based multi-phase study to validate the measurement
18
19 169 properties of a new PROM, the LEQ-CI. The planned study will take place over a 12-month
20
21 170 period and has been co-produced by the study team with input from two lay members, both CI
22
23 171 recipients, from the study's Research Management Group. Lay members reviewed and
24
25 172 provided feedback on study design, participant documents and iterations of the LEQ-CI.
26
27
28
29 173

30 31 174 **Development of the LEQ-CI**

32
33 175 The LEQ-CI is a hearing-specific PROM measuring listening effort in adult cochlear
34
35 176 implant candidates and recipients. It is comprised of 29 items across four domains. Five-point
36
37 177 or seven-point Likert scales with absolute anchors and labelled categories ensure a broad range
38
39 178 of response options. Item responses are summed to produce a simple total score. The LEQ-CI's
40
41 179 conceptual framework, presented in Figure 1, was developed from a mixed-methods qualitative
42
43 180 study involving focus groups and a postal survey.[2] An item bank was constructed that
44
45 181 included new items and items harvested from extant PROMs considered to measure listening
46
47 182 effort or associated constructs.[17,18] Exemplar items are presented in Figure 2. Preliminary
48
49 183 testing was completed to identify and rectify problems with items and response scales prior to
50
51 184 undertaking psychometric evaluation.[31] Preliminary testing involved the use of multiple
52
53 185 datasets to assess reliability.[32] Item quality was estimated using the on-line Survey Quality
54
55 186 Predictor system (SQP 2.1, <http://sqp.upf.edu/>). An expert review panel of academics,
56
57
58
59
60

1
2
3 187 researchers, and clinicians (n = 7) and a series of cognitive interviews with a purposive sample
4
5 188 of cochlear implant candidates and recipients (n = 12) were completed to elicit feedback on the
6
7
8 189 relevance, clarity and acceptability of the LEQ-CI.
9

10 190

11 12 191 **Sample size**

13
14 192 The study sample will be representative of the population of adults with acquired, post-
15
16 193 lingual severe-to-profound sensorineural hearing loss (SNHL) referred for cochlear
17
18 194 implantation in the UK. A total study sample of 350 participants will be recruited. In Phase 1,
19
20 195 a cohort of 250 participants will be recruited from four National Health Service (NHS) cochlear
21
22 196 implant centres. To minimise burden on implant centre staff and to ensure representation from
23
24 197 different regions of the UK, each centre will send questionnaire packs to 125 cochlear implant
25
26 198 candidates or recipients who meet the study inclusion criteria (n = 500). If necessary, additional
27
28 199 participants will be recruited until such time as 250 completed LEQ-CI forms with no missing
29
30 200 data are returned. There are no general criteria for determination of sample sizes in studies of
31
32 201 PROM validation and sample sizes are, in part, dependent on the psychometric characteristics
33
34 202 being assessed.[23,27] Mokkink et al. recommend greater than 200 respondents when
35
36 203 undertaking RA and seven times the number of items for purposes of undertaking assessment
37
38 204 of unidimensionality.[23] Linacre recommends a minimum sample size of 250 respondents for
39
40 205 definitive item calibration using RA.[33] The LEQ-CI is comprised of 29 items that have been
41
42 206 selected to minimise respondent burden whilst allowing for adequate sampling of relevant
43
44 207 constructs associated with listening effort. Therefore, a minimum sample of 250 participants
45
46 208 is considered sufficient for undertaking both assessment of unidimensionality and RA of the
47
48 209 LEQ-CI. In Phase 2, a new cohort of 100 participants fulfilling the same eligibility criteria will
49
50 210 be recruited from two cochlear implant centres. Each centre will recruit 125 participants
51
52 211 initially. If necessary, further participants will be recruited until the required sample size is
53
54
55
56
57
58
59
60

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

214

215 **Recruitment and data collection**

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

218 Table 1. Study eligibility criteria for recruitment of participants

Inclusion Criteria	Exclusion Criteria
<ul style="list-style-type: none"> • Adults (persons \geq 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 	<ul style="list-style-type: none"> • 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).

219

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

1
2
3 224 presumed if the questionnaires are completed and returned to the study team. To maintain
4
5 225 participant anonymity, eligibility screening and the study documents will be mailed to
6
7 226 prospective participants by a member of the clinical team at each participating cochlear implant
8
9 227 centre. To maintain participant anonymity, each questionnaire pack will be coded with a unique
10
11 228 identifier. No personal identifiable information will be retained by the study team.
12
13
14
15 229

16 17 230 **Statistical Analysis**

18
19 231 There are two schools of psychometric measurement theory dominate the field of
20
21 232 PROM development.[29,36] Traditional psychometric analyses (e.g., Cronbach's alpha as a
22
23 233 measure of internal consistency reliability) are underpinned by CTT. CTT seeks to evaluate
24
25 234 reliability and validity of a scale and has been the dominant approach used in the development
26
27 235 and validation of outcome measures.[37] However, modern measurement techniques such as
28
29 236 RA are increasingly being reported alongside traditional analyses in studies of PROM
30
31 237 development and validation (e.g., [38,39]).
32
33
34

35 238 CTT is based on the assumption that every observed score is a function of an
36
37 239 individual's true score and random error.[40] The assumptions underpinning CTT differ from
38
39 240 those underpinning the Rasch model. It has been argued that CTT cannot be adequately be
40
41 241 tested as it is based on definitions rather than assumptions which can be proven true or false.
42
43 242 This is in contrast to modern measurement theory (i.e., RA) which can generate assumptions
44
45 243 that can be proven true or false.[41] Whereas CTT methods focus on the total score of a
46
47 244 measure, RA enables instrument developers to focus more specifically on the characteristics of
48
49 245 individual items.[42] For example, RA, unlike CTT methods, can be used to establish whether
50
51 246 an item's response scale is functioning as expected and, if not, suggest improvements.
52
53
54

55
56 247 The Rasch model allows for ordering persons (i.e., patients) according to the amount
57
58 248 of the latent target construct (i.e., listening effort) they possess and for ordering items that
59
60

1
2
3 249 measure the target construct according to their difficulty.[36] This method allows non-linear
4
5 250 (i.e., ordinal) raw data to be converted to a linear (i.e., interval) scale, which can then be
6
7
8 251 evaluated through the use of parametric statistical tests.[43] By contrast, CTT methods yield
9
10 252 measures that produce ordinal rather than interval level data. This has implications for the
11
12 253 interpretation of test scores as difference scores and changes scores are most meaningful when
13
14 254 interval level of measurement is used.[36,41]

15
16
17 255 A further limitation of CTT is that the performance of a test is dependent on the sample
18
19 256 in which that test is assessed.[41] This renders its psychometric properties (i.e., reliability and
20
21 257 validity) dependent on the sample rather than characteristics of the test itself. By contrast, RA
22
23 258 produces item and test statistics that are sample independent rendering the test valid across
24
25 259 groups. Any discrepancies between the scale data and the Rasch model requirements are
26
27 260 indicative of anomalies in the scale as a measurement instrument. These discrepancies provide
28
29 261 diagnostic information that serves as a basis for understanding and empirical improvement of
30
31 262 the instrument at both item and scale-level.[44]

32
33
34
35 263 Despite these limitations, CTT methods continue to be widely used in studies of
36
37 264 instrument validation and are included in the COSMIN standards.[28,41] Indeed, some
38
39 265 properties (e.g., acceptability, scaling assumptions) can only be evaluated using CTT
40
41 266 methods.[37] For these reasons, this study will use both CTT and RA in a complementary
42
43 267 fashion to ensure rigorous validation of the LEQ-CI at both item and scale level.

44
45
46 268 Study data will be managed using the online clinical data management programme,
47
48 269 REDCap (Version 7.2.1, Vanderbilt University), licensed to the Swansea Trials Unit, Swansea
49
50 270 University. RA will be used to evaluate the LEQ-CI's psychometric properties and refine the
51
52 271 item and scale structure of the LEQ-CI using Winsteps (Version 4.1.0) software. Psychometric
53
54 272 analyses applying CTT will be conducted using the Statistical Package for Social Sciences
55
56 273 (SPSS), Version 22.0 licensed to Swansea University.

274

275 Phase 1: Item & Scale Refinement using Rasch Measurement Theory

276 *Assessing unidimensionality*

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

281 *Assessing item fit*

282 In RA, item fit refers to the degree of mismatch between the pattern of actual observed
283 responses and the Rasch modelled expectations. Specifically, it refers to the pattern for each
284 item across persons investigated by examining item infit and outfit statistics.[36] Mean square
285 standardized residuals (MNSQ) will be used to assess fit with MNSQ residuals within the 0.5–
286 1.5 range considered acceptable for productive measurement. Mean square values less than 0.5
287 indicate overfit (i.e., the items are too predictable relative to the Rasch model), while mean
288 square values greater than 1.5 are indicative of too much noise (randomness) relative to the
289 Rasch model.[47]

290 *Assessing differential item functioning (DIF)*

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

1
2
3 299 DIF they will either be considered candidates for removal or examined for adjustment of DIF
4
5 300 and re-evaluated, thus reflecting the iterative nature of instrument validation.[36,49]
6

7
8 301 *Assessing response scale ordering*
9

10 302 The response options of an instrument (i.e., number of categories and their definitions)
11
12 303 are critical to its reliability and validity.[50] The Rasch model will enable us to show
13
14 304 empirically how respondents use the LEQ-CI's rating scale informing future iterations of the
15
16 305 LEQ-CI to ensure it yields high quality data.[36] Response category ordering will be assessed
17
18 306 using Rasch probability curves and there will be an examination of the data for category
19
20 307 disordering and threshold disordering.[51] These investigations will show whether the
21
22 308 response options selected for the LEQ-CI are sufficient or should be collapsed to provide better
23
24 309 coverage of the latent trait.
25
26
27

28 310 *Assessing the targeting of persons and items*
29

30 311 Targeting using RA explores whether the instrument has a distribution of items that
31
32 312 matches the range of the respondents' latent trait. This will be done by examining the item-
33
34 313 person threshold distribution map, which illustrates a relative position of "item difficulty" to
35
36 314 "person ability".[45] The means and standard deviations of items and persons should match
37
38 315 closely.[36]
39
40
41

42 316 *Assessing reliability*
43

44 317 The reliability of the LEQ-CI will be examined by observing the person separation
45
46 318 index (PSI). The PSI is an estimate of the spread or separation of persons on the measured
47
48 319 variable [36] and is considered to be a measure of internal consistency reliability.[52] It is a
49
50 320 measure of the scale's ability to separate the study sample. A $PSI > 0.7$ will be considered an
51
52 321 adequate measure of reliability.[53]
53
54
55

56 322

57
58 323 Phase 1: Psychometric evaluation using Classical Test Theory
59
60

1
2
3 324 *Assessing acceptability and data completeness*
4

5 325 Acceptability and data completeness will establish extent to which scale items are
6
7
8 326 scored and total scores can be computed. Assessment of the completeness of item and scale-
9
10 327 level data (i.e., missing or incomplete data for items and sample) including frequency of
11
12 328 endorsement will be completed. Score distributions including skew of scale scores and
13
14 329 presence of floor and ceiling effects will be examined.[29]

16
17 330 *Assessing scaling assumptions*
18

19 331 Examination of scaling assumptions involves assessment of whether it is legitimate to
20
21 332 group items into a scale to produce a scale score. Tests of scaling assumptions examine item-
22
23 333 total correlations, mean scores and standard deviations. When checking homogeneity of the
24
25 334 LEQ-CI's scales, the heuristic that items should correlate with the total score above 0.20 will
26
27 335 be applied. Item-total correlations will be calculated using the Pearson product-moment
28
29 336 correlation.[29]

30
31
32
33 337 *Assessing targeting*
34

35 338 Targeting may be defined as “the extent to which the range of the variable measured by
36
37 339 the scale matches the range of that variable in the study sample” (p.4).[54] Targeting will be
38
39 340 assessed following item refinement using RA. CTT will be used determine whether the LEQ-
40
41 341 CI scale scores span the entire scale range, skewness, and whether floor and ceiling effects are
42
43 342 low, defined as < 15% of the sample.[27]

44
45
46
47 343 *Assessing internal consistency reliability*
48

49 344 Assessment of internal consistency establishes the inter-relatedness among items and
50
51 345 is an assessment of the unidimensionality of a scale or subscale.[23] Internal consistency will
52
53 346 be assessed by calculating inter-item and item-total correlations and Cronbach's alpha.

- 54
55
56 347 • Inter-item correlation – Calculating inter-item correlations will provide an indication
57
58 348 whether an item is part of a (sub)scale. Correlations should fall between 0.2 and 0.5. Items
59
60

1
2
3 349 which have a correlation greater than 0.7 may be considered to measure the same thing,
4
5 350 making one item a candidate for deletion.[27]
6

7
8 351 • Item-total correlation – Calculating item-total correlations will assess whether the LEQ-
9
10 352 CI's items discriminate patients on the listening effort construct. Items that show an item-
11
12 353 total correlation of less than 0.3 will be considered as contributing little to the LEQ-CI in
13
14 354 terms of discriminating between individuals with high versus low levels of listening
15
16 355 effort.[27] These items will be considered candidates for deletion.
17
18

19 356 • Cronbach's alpha - Internal consistency will be calculated for each subscale of the LEQ-CI
20
21 357 by calculating Cronbach's alpha. Alpha values ≥ 0.70 and ≤ 0.95 will be considered
22
23 358 good evidence of internal consistency.[29]
24
25

26 359

27 28 29 360 Phase 2: Establishing construct validity

30
31 361 Construct validity may be defined as the extent to which the scores of an instrument are
32
33 362 a valid measure of the latent construct.[27] Construct validity of the refined LEQ-CI will be
34
35 363 assessed by applying criteria specified by the COSMIN group. COSMIN guidance specifies
36
37 364 that construct validity may be assessed by testing a priori hypotheses based on the literature
38
39 365 and the experience of the study team.[23] Hypotheses are generated by the study team and
40
41 366 founded on the assumption that the LEQ-CI validly measures the target construct (i.e., listening
42
43 367 effort). These state the relationship between the instrument and other measures, as well as the
44
45 368 expected differences between the scores attained by different sub-groups of the target
46
47 369 population.. To establish the construct validity of an instrument, Mokkink et al. recommend at
48
49 370 least 75% of the stated hypotheses are endorsed.[23]
50
51

52 371 *Assessing convergent validity*

53
54 372 Concurrent construct validity will be assessed by examining the correlation between
55
56 373 scores on the LEQ-CI with the summed score on the three items considered to measure listening
57
58
59
60

1
2
3 374 effort on the Speech, Spatial, and Qualities of Hearing questionnaire (SSQ). As no validated
4
5 375 measure of listening effort has been identified as a suitable comparator PROM these items were
6
7
8 376 selected to assess construct validity as the SSQ has good evidence of being a well-validated
9
10 377 instrument across multiple studies.[17] As the LEQ-CI and the SSQ are measuring the same
11
12 378 construct, we hypothesise that a strong positive correlation > 0.50 will be observed between
13
14 379 measures as suggested by Mokkink et al.[23] We further predict a moderate positive correlation
15
16
17 380 (0.30 – 0.50) between the LEQ-CI and SSQ total score as LE may be considered to be a
18
19 381 component of hearing disability, the construct measured by the SSQ.[11]

21 382 *Assessing discriminant validity*

23
24 383 Discriminant (i.e., divergent) validity is an assessment of a measure's ability to
25
26 384 discriminate between dissimilar constructs.[29] It will be assessed by the examining the
27
28 385 correlation between scores on the LEQ-CI and the Fatigue Assessment Scale (FAS),[55] a
29
30 386 measure of fatigue used in other studies investigating LE and fatigue in individuals with
31
32
33 387 HL.[56] As LE and fatigue are similar but unrelated constructs [16] we anticipate a moderate
34
35 388 positive correlation between 0.30 – 0.50.

37
38 389 Further assessment of discriminant validity will be undertaken by examining the
39
40 390 correlation between scores on the LEQ-CI and the Nijmegen Cochlear Implant Questionnaire
41
42 391 (NCIQ),[57] a measure of quality of life in CI patients. A small positive correlation of between
43
44 392 0.30 – 0.50 is anticipated as these measures may be considered to assess similar, but unrelated
45
46
47 393 constructs.[23]

48
49 394

51 395 **ETHICS AND DISSEMINATION**

53
54 396 This study has received ethical approval from the NHS Research Ethics Committee
55
56 397 (REC): Newcastle and North Tyneside 2 (Ref: 18/NE/0320). Study findings will be
57
58
59
60

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

400

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

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

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

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

408 **Open Access:** This is an Open Access article distributed in accordance with the terms of the
409 Creative Commons Attribution (CC BY NC 4.0) license, which permits others to distribute,
410 remix, adapt and build upon this work, for non-commercial use, provided the original article is
411 properly cited. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

413

414

REFERENCES

- 415 1 Wilson BS, Tucci DL, Merson MH, *et al.* Global hearing health care: new findings and
416 perspectives. *Lancet* 2017;**390**:2503–15.
- 417 2 Hughes SE, Hutchings HA, Rapport FL, *et al.* Social connectedness and perceived
418 listening effort in adult cochlear implant users: a Grounded Theory to establish content
419 validity for a new patient reported outcome measure. *Ear Hear* 2018;**39**:922-34.
- 420 3 McGarrigle R, Munro K, Dawes, *et al.* Listening effort and fatigue: what exactly are
421 we measuring? A British Society of Audiology Cognition in Hearing Special Interest
422 Group 'white paper'. *Int J Audiol* 2014;**53**:433-40.

- 1
2
3 423 4 Mackersie CL, MacPhee IX, Heldt EW. Effects of hearing loss on heart rate variability
4
5 424 and skin conductance measured during sentence recognition in noise. *Ear Hear*
6
7 425 2015;**36**:145-54..
8
9
10 426 5 Zekveld AA, George ELJ, Houtgast T, *et al.* Cognitive abilities relate to self-reported
11
12 427 hearing disability. *J Speech Lang Hear Res* 2013;**56**:1364–72.
13
14 428 6 Ramage-Morin PL. Hearing difficulties and feelings of social isolation among
15
16 429 Canadians aged 45 or older. 2016;**27**:3-12.[http://www.statcan.gc.ca/pub/82-003-](http://www.statcan.gc.ca/pub/82-003-x/2016011/article/14671-eng.pdf)
17
18 430 [x/2016011/article/14671-eng.pdf](http://www.statcan.gc.ca/pub/82-003-x/2016011/article/14671-eng.pdf) (accessed 23 Apr 2017).
19
20
21 431 7 Nachtegaal J, Kuik DJ, Anema JR, *et al.* Hearing status, need for recovery after work,
22
23 432 and psychosocial work characteristics: results from an internet-based national survey
24
25 433 on hearing. *Int J Audiol* 2009;**48**:684–91.
26
27
28 434 8 Hutchings HA, Alrubaiy L. Patient-Reported Outcome Measures in Routine Clinical
29
30 435 Care : The PROMise of a Better Future? *Dig Dis Sci.*2017;**62**:1841-43.
31
32
33 436 9 Ousey K, Cook L. Understanding patient reported outcome measures (PROMs). *Br J*
34
35 437 *Community Nurs* 2011;**16**:80–2.
36
37
38 438 10 Meadows KA. Patient-reported outcome measures: an overview. *Br J Community Nurs*
39
40 439 2011;**16**:146–51.
41
42 440 11 Noble W. Self-Assessment of Hearing. 2nd ed. San Diego:Plural Publishing 2013.
43
44 441 12 Devlin NJ, Appleby J. Getting the most out of PROMs: putting health outcomes at the
45
46 442 heart of NHS decision-making. London: 2010.
47
48 443 [http://www.kingsfund.org.uk/sites/files/kf/Getting-the-most-out-of-PROMs-Nancy-](http://www.kingsfund.org.uk/sites/files/kf/Getting-the-most-out-of-PROMs-Nancy-Devlin-John-Appleby-Kings-Fund-March-2010.pdf)
49
50 444 [Devlin-John-Appleby-Kings-Fund-March-2010.pdf](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).
51
52
53 445 13 Hornsby BWY. The effects of hearing aid use on listening effort and mental fatigue
54
55 446 associated with sustained speech processing demands. *Ear Hear* 2013;**34**:523–34.
56
57
58 447 14 Alhanbali S, Dawes P, Millman RE, *et al.* Measures of listening effort are
59
60

- 1
2
3 448 multidimensional. *Ear Hear* 2019;e-pub ahead of print:1–14.
- 4
5 449 15 Miles K, McMahon C, Boisvert I, *et al.* Objective assessment of listening effort:
6
7
8 450 coregistration of pupillometry and EEG. *Trends Hear* 2017;**21**:1-13.
- 9
10 451 16 Alhanbali S, Dawes P, Lloyd S, *et al.* Hearing handicap and speech recognition
11
12 452 correlate with self-reported listening effort and fatigue. *Ear Hear* 2018;**39**:470-74.
- 13
14 453 17 Hughes SE, Hutchings HA, Dobbs TD, *et al.* A systematic review and narrative
15
16 454 synthesis of the measurement properties of patient-reported outcome measures
17
18 455 (PROMs) used to assess listening effort in hearing loss [poster presentation]. British
19
20 456 Society of Audiology 2017;Harrogate.
- 21
22 457 18 Hughes SE, Rapport FL, Boisvert I, *et al.* Patient-reported outcome measures
23
24 458 (PROMs) for assessing perceived listening effort in hearing loss: protocol for a
25
26 459 systematic review. *BMJ Open* 2017 June 6;**7**(5):e014995.
- 27
28 460 19 Gatehouse S, Noble W. The speech, spatial and qualities of hearing scale (SSQ). *Int J*
29
30 461 *Audiol* 2004;**43**:85–99.
- 31
32 462 20 Cox RM, Gilmore C, Alexander GC. Comparison of two questionnaires for patient-
33
34 463 assessed hearing aid benefit. *J Am Acad Audiol* 1991;**2**:134–45.
- 35
36 464 21 Cox RM, Alexander GC. The Abbreviated Profile of Hearing Aid Benefit. *Ear Hear*
37
38 465 1995;**16**:176–86.
- 39
40 466 22 Demorest ME, Erdman SUEANN. Development of the communication profile for the
41
42 467 hearing impaired. *J Speech Hear Disord* 1987;**52**:129–43.
- 43
44 468 23 Mokka LB, Prinsen CAC, Patrick DL, *et al.* COSMIN Methodology for Systematic
45
46 469 Reviews of Patient - Reported Outcome Measures (PROMs) User Manual Version 1.0.
47
48 470 2018 February:1–78. [www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-](http://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf)
49
50 471 [PROMs-manual_version-1_feb-2018-1.pdf](http://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf) (accessed 4 Apr 2019)
- 51
52 472 24 Strauss DJ, Francis AL. Toward a taxonomic model of attention in effortful listening.
53
54
55
56
57
58
59
60

- 1
2
3 473 *Cogn Affect Behav Neurosci* 2017;**17**:809–25.
4
5 474 25 Peelle JE. Listening effort: how the cognitive consequences of acoustic challenge are
6 reflected in brain and behavior. *Ear Hear* 2018;**39**:204-14.
7
8 475
9
10 476 26 Pichora-Fuller MK, Kramer SE, Eckert MA, *et al.* Hearing impairment and cognitive
11 energy. *Ear Hear* 2016;**37**:5S-27S.
12 477
13
14 478 27 de Vet H, Terwee CB, Mokkink LB, *et al.* Measurement in Medicine: A Practical
15 Guide. Cambridge:Cambridge University Press 2011.
16 479
17
18 480 28 Mokkink LB, de Vet HCW, Prinsen CAC, *et al.* COSMIN Risk of bias checklist for
19 systematic reviews of patient-reported outcome measures. *Qual Life Res*
20 481
21 2018;**27**:1171-79.
22 482
23
24 483 29 Streiner DL, Norman GR. Health Measurement Scales: A Practical Guide to Their
25 Development and Use. 4th ed. Oxford University Press 2008.
26 484
27
28 485 30 Prinsen CAC, Mokkink LB, Bouter LM, *et al.* COSMIN guideline for systematic
29 reviews of patient-reported outcome measures. *Qual Life Res* 2018;**27**:1147–57.
30 486
31
32 487 31 Hughes SE, Boisvert I, Rapport F, *et al.* Measuring the experience of listening effort in
33 cochlear implant candidates and recipients: Development and pretesting of the LEQ-CI
34 [poster presentation]. Southampton: British Cochlear Implant Group Academic
35 488
36 Meeting 2019.
37 489
38
39 490
40
41 491 32 Maitland A, Presser S. How accurately do different evaluation methods predict the
42 reliability of survey questions? *J Surv Stat Methodol* 2016;**4**:362-81.
43 492
44
45 493 33 Linacre JM. Sample size and item calibration or person measure stability. *Rasch Meas*
46 *Trans* 1994;**7**:328.
47 494
48
49 495 34 Hobart JC, Cano SJ, Warner TT, *et al.* What sample sizes for reliability and validity
50 studies in neurology? *J Neurol* 2012;**259**:2681–94.
51 496
52
53 497 35 National Institute for Health and Clinical Excellence (NICE). NICE technology
54
55
56
57
58
59
60

- 1
2
3 498 appraisal guidance 166: Cochlear implants for children and adults with severe to
4
5 499 profound deafness. 2009. www.guidance.nice.org.uk/TA166/pdf/English (accessed
6
7 500 Dec 2018).
- 8
9
10 501 36 Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the
11
12 502 Human Sciences 3rd ed. London:Routledge 2015.
- 13
14 503 37 Petrillo J, Cano SJ, Mcleod LD, *et al.* Using Classical Test Theory, Item Response
15
16 504 Theory, and Rasch Measurement Theory to evaluate patient-reported outcome
17
18 505 measures: a comparison of worked examples. *Value Heal* 2015;**18**:25–34.
- 19
20 506 38 Barker BA, Donovan NJ, Schubert AD, *et al.* Using Rasch analysis to examine the
21
22 507 item-level psychometrics of the *Infant-Toddler Meaningful Auditory Integration*
23
24 508 *Scales*. *Speech, Lang Hear* 2017;**20**:130–43.
- 25
26 509 39 Heffernan E, Maidment DW, Barry JG, *et al.* Refinement and validation of the Social
27
28 510 Participation Restrictions Questionnaire: an application of Rasch analysis and
29
30 511 traditional psychometric analysis techniques. *Ear Hear* 2018;**40**:328-39.
- 31
32 512 40 Tractenberg RE. Classical and modern measurement theories, patient reports and
33
34 513 clinical outcomes. *Comtemp Clin Trials* 2010;**31**:1–3.
- 35
36 514 41 Cano SJ, Hobart JC, Cano S. The problem with health measurement. *Patient Prefer*
37
38 515 *Adherence* 2011;**5**:279-90.
- 39
40 516 42 da Rocha, NS, Chachamovich E, de Almeida Fleck, MP *et al.* An introduction to
41
42 517 Rasch analysis for psychiatric practice and research. *J Psychiatr Res* 2013;**47**:141–8.
- 43
44 518 43 Boone WJ. Rasch analysis for instrument development: why, when, and how? *CBE*
45
46 519 *Life Sci Educ* 2016;**15**: rm4.
- 47
48 520 44 Ball S, Vickery J, Hobart J, *et al.* The Cannabinoid Use in Progressive Inflammatory
49
50 521 brain Disease (CUPID) trial: a randomised double-blind placebo-controlled parallel-
51
52 522 group multicentre trial and economic evaluation of cannabinoids to slow progression
53
54
55
56
57
58
59
60

- 1
2
3 523 in multiple sclerosis. *Health Technol Assess* 2015;**19**:1–187.
4
5 524 45 Tadic V, Tadi V, Cooper A, *et al.* Measuring the quality of life of visually impaired
6
7 525 children: first stage psychometric evaluation of the novel VQoL-CYP. *PLoS One*
8
9 526 2016;**11**:e0146225.
10
11
12 527 46 Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical
13
14 528 assessment instruments. *Psychol. Assess* 1995;**7**:286–99.
15
16
17 529 47 Wright, Linacre. Reasonable mean-square fit vales. *Rasch Meas Trans*
18
19 530 1994;**8**.www.rasch.org (accessed 20 Feb 2018).
20
21
22 531 48 Zumbo B. A handbook on the theory and methods of differential item functioning
23
24 532 (DIF): logistic regression modeling as a unitary framework for binary and likert-type
25
26 533 (ordinal) item scores. *Directorate of Human Resources Research and Evaluation,*
27
28 534 *Department of National Defense*
29
30 535 1999.[www.researchgate.net/publication/236596822_A_handbook_on_the_theory_and](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)
31
32 536 [_methods_of_differential_item_functioning_\(DIF\)_Logistic_regression_modeling_as_](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)
33
34 537 [a_unitary_framework_for_binary_and_Likert-](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)
35
36 538 [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).
37
38
39 539 49 Teresi JA, Ramirez M, Jones RN, *et al.* Modifying measures based on differential item
40
41 540 functioning (DIF) impact analyses. *J Aging Health* 2012;**24**:1044–76.
42
43
44 541 50 Johnston M, Dixon D, Hart J, *et al.* Discriminant content validity: a quantitative
45
46 542 methodology for assessing content of theory-based measures, with illustrative
47
48 543 applications. *Br J Health Psychol* 2014;**19**:240–57.
49
50
51 544 51 Linacre JM. Category Disordering (disordered categories) vs. Threshold Disordering
52
53 545 (disordered thresholds). *Rasch Meas Trans* 1999;**13**:675.
54
55
56 546 52 Gorecki CA. The development and validation of a patient-reported outcome measure
57
58 547 of health-related quality of life for patients with pressure ulcers: PUQOL Project.
59
60

- 1
2
3 548 University of Leeds 2011.
4
5
6 549 53 Tennant, Conaghan. The measurement model model in rheumatology: what is it and
7
8 550 why use it? When should it be applied, and what should one look for in a Rasch paper?
9
10 551 *Arthritis Care Res* 2007;**57**:1358-62
11
12 552 54 Gorecki C, Brown JM, Cano S, *et al.* Development and validation of a new patient-
13
14 553 reported outcome measure for patients with pressure ulcers: the PU-QOL instrument.
15
16 554 *Health Qual Life Outcomes* 2013;**11**:95.
17
18
19 555 55 Michielsen HJ, De Vries J, Van Heck GL, *et al.* Examination of the dimensionality of
20
21 556 fatigue. *Eur J Psychol Assess* 2004;**20**:39–48.
22
23
24 557 56 Alhanbali S, Dawes P, Lloyd S, *et al.* Self-reported listening-related effort and fatigue
25
26 558 in hearing-impaired adults. *Ear Hear* 2017;**38**:e39–48.
27
28 559 57 Hinderink JB, Krabbe PF, Van Den Broek P. Development and application of a health-
29
30 560 related quality-of-life instrument for adults with cochlear implants: the Nijmegen
31
32 561 Cochlear Implant Questionnaire. *Otolaryngol Head Neck Surg* 2000;**123**:756–65.
33
34
35
36
37
38 563
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

564

FIGURE 1

565 The conceptual frame work of the LEQ-CI

566

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

567

FIGURE 2

568 Example items from the LEQ-CI

569

For peer review only

Figure 1. The conceptual framework of the LEQ-CI

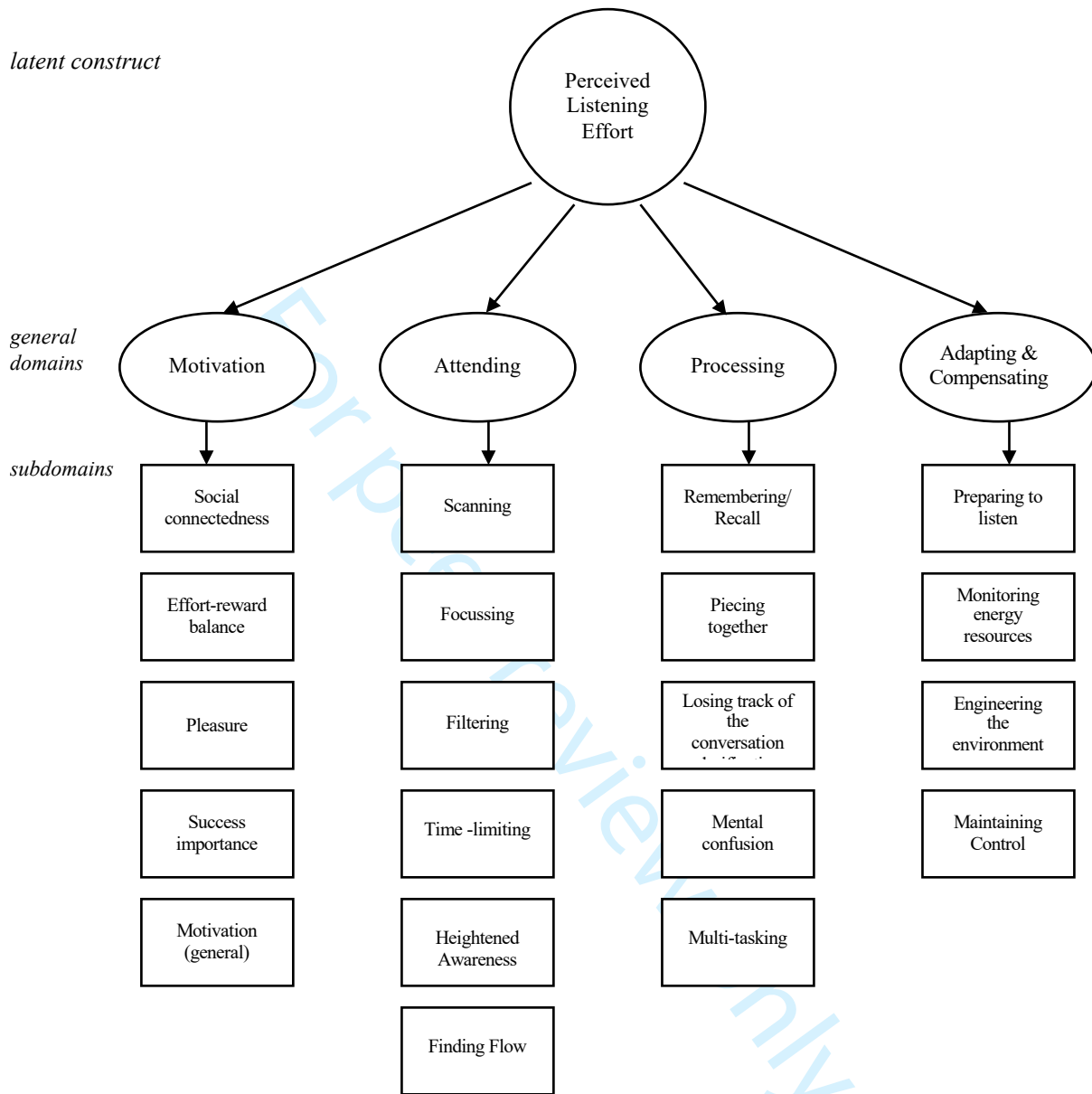


Figure 2: Example items from the LEQ-CI

Domain	Item
Attending	<p>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></p> <p>Response Scale: 7-point Likert scale (1 = never; 2 = rarely; 3 = occasionally; 4 = sometimes; 5 = frequently; 6 = usually; 7 = always)</p>
Processing	<p>E323 In a typical week, are you able to listen to someone talk while doing something else? <i>Please tick one box.</i></p> <p>Response Scale: 7-point Likert scale (1 = never; 2 = rarely; 3 = occasionally; 4 = sometimes; 5 = frequently; 6 = usually; 7 = always)</p>
Adapting & compensating	<p>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></p> <p>Response Scale: 5-point Likert scale (1 = never; 2 = 1-2 days; 3 = 3-4 days; 4 = 5-6 days; 5 = Everyday)</p>
Motivation	<p>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></p> <p>Response Scale: 7-point Likert scale (1 = never; 2 = rarely; 3 = occasionally; 4 = sometimes; 5 = frequently; 6 = usually; 7 = always)</p>