

Title: Development of the Social Participation Restrictions Questionnaire (SPaRQ) through consultation with adults with hearing loss, researchers, and clinicians: A content evaluation study

Authors: Eithne Heffernan^{a,b,*}, Neil S Coulson^c, Melanie A Ferguson^{a,b,d}

^aNational Institute of Health Research Nottingham Biomedical Research Centre, Nottingham, United Kingdom

^bHearing Sciences Group, Division of Clinical Neuroscience, School of Medicine, University of Nottingham, Nottingham, United Kingdom

^cDivision of Rehabilitation and Ageing, School of Medicine, University of Nottingham, Nottingham, United Kingdom

^dNottingham University Hospitals NHS Trust, Nottingham, United Kingdom

***Corresponding author:** NIHR Nottingham Biomedical Research Centre, Ropewalk House, 113 The Ropewalk, Nottingham, United Kingdom, NG1 5DU.

Tel: +44115 8232631

Email: eithne.heffernan1@nottingham.ac.uk

Abbreviations:

AHL – Adult with hearing loss

BRC – Biomedical Research Centre

ICF – International Classification of Functioning, Disability, and Health

NIHR – National Institute for Health Research

PPI – Patient and Public Involvement

PROM – Patient-reported outcome measure

SME – Subject matter expert

SPaRQ – Social Participation Restrictions Questionnaire

WHO – World Health Organization

30 **Abstract**

31 **Objective** - This research aimed to evaluate the content of the Social Participation
32 Restrictions Questionnaire (SPaRQ) in terms of its relevance, clarity, comprehensiveness,
33 acceptability to adults with hearing loss, and responsiveness.

34 **Design** - Cognitive interviews and a subject matter expert survey were conducted. The
35 interview data were analysed using thematic analysis and a taxonomy of questionnaire clarity
36 problems. Descriptive statistics were calculated for the survey data.

37 **Study sample** - Fourteen adults with hearing loss participated in the cognitive interviews.
38 Twenty clinicians and academics completed the subject matter expert survey.

39 **Results** - The majority of the SPaRQ content was found to be relevant, clear, comprehensive,
40 and acceptable. However, an important clarity problem was identified: many adults with
41 hearing loss struggled to switch from answering positively-worded items (e.g. 'I can attend
42 social gatherings') to answering negatively-worded items (e.g. 'I feel isolated'). Several
43 subject matter experts found responsiveness difficult to assess. The SPaRQ was amended
44 where necessary.

45 **Conclusions** - Few hearing-specific questionnaires have undergone content evaluation. This
46 study highlights the value of content evaluation as a means of identifying important flaws and
47 improving the quality of a measure. The next stage of this research is a psychometric
48 evaluation of the measure.

49 **Introduction**

50 Participation restrictions have been defined as the difficulties an individual experiences with
51 involvement in life situations (World Health Organization 2001). These situations include
52 family relationships, friendships, recreation, community life, education, and employment
53 (Danermark et al. 2013). Numerous studies have demonstrated that participation restrictions
54 are one of the major negative consequences of hearing loss (Barker et al. 2016; Vas et al.
55 2017). Therefore, one of the main aims of auditory rehabilitation is to reduce participation
56 restrictions in individuals with hearing loss (Boothroyd 2007; Ferguson et al. 2017; Ferguson
57 et al. in press).

58 In order to evaluate the impact of auditory rehabilitation on participation restrictions,
59 it is necessary to have a valid, hearing-specific outcome measure for this construct. However,
60 participation restrictions are recognised as being one of the most difficult constructs to
61 measure (Salter et al. 2005; Whiteneck and Dijkers 2009). Much of this difficulty stems from
62 the broad and inconsistent conceptualisation of the construct (Heinemann et al. 2010). The
63 World Health Organization's (2001) definition of participation restrictions does not readily
64 lend itself to measurement, as 'life situations' could refer to practically any situation between
65 birth and death (Dijkers 2010). There is little agreement in the literature concerning the
66 domains (e.g. communication, community life) that should be included in a participation
67 restrictions measure. Furthermore, it has proven difficult to distinguish participation
68 restrictions from similar constructs, such as quality of life, activity, and social support
69 (Whiteneck and Dijkers 2009; Eyssen et al. 2011).

70 It is clear that, in order to develop a valid measure of participation restrictions, it is
71 first necessary to develop a strong conceptual foundation for that measure. This can be
72 achieved by following best practice recommendations from the questionnaire development

literature (e.g. Brod et al. 2009; Mokkink et al. 2012; Reeve et al. 2013). Specifically, it is recommended that questionnaire developers conceptualise the target construct through an in-depth literature review and qualitative research with key stakeholders (e.g. patients and clinicians). The findings are used to generate a clear definition and conceptual model of the target construct. The conceptual model comprises domains and subdomains that serve as the basis of the subscales and items of the measure. Ideally, the words and phrases used by the patients should be incorporated in the items (Haynes et al. 1995; Rattray and Jones 2007; Brod et al. 2009).

Once a prototype of the measure has been created, it is recommended that it is thoroughly reviewed by key stakeholders to identify and rectify any flaws in its design (McGartland Rubio et al. 2003; Brod et al. 2009). For instance, the measure could omit important content or include unimportant content, which degrades the quality of the data collected via that measure, as well as the quality of the clinical inferences drawn from those data (Haynes et al. 1995). Furthermore, aspects of the questionnaire could be difficult to understand, especially abstract expressions and technical terms. Such problems are difficult to detect without stakeholder feedback, as some respondents answer items that they do not understand out of a sense of politeness or duty, whilst others answer items ‘mindlessly’ without realising that they have misunderstood them (Collins 2003). Additionally, respondents could differ in their interpretations of the questionnaire. For example, respondents can have different interpretations of seemingly unambiguous terms such as ‘Always’ and ‘Never’ (Aronson 2006). Finally, the questionnaire could include terms that are offensive or off-putting to particular cultural groups (Boynton et al. 2004). Therefore, stakeholder feedback is vital to ensuring that valuable resources are not wasted by utilising a questionnaire that is inherently flawed in a large quantitative study (McGartland Rubio et al.

2003). Despite this, qualitative research with key stakeholders has seldom been used in the development of hearing-specific questionnaires.

This research set out to develop a high-quality, hearing-specific measure of participation restrictions in accordance with best practice recommendations (Brod et al. 2009; Mokkink et al. 2012). This questionnaire, entitled the Social Participation Restrictions Questionnaire (SPaRQ), was specifically designed to be a standardised, self-administered, patient-reported outcome measure (PROM) for use in research and practice with adults with hearing loss. The development of the initial prototype of the SPaRQ began with the conceptualisation of the target construct: hearing-related participation restrictions. This was achieved by reviewing:

- 1) The findings of semi-structured interviews with adults with hearing loss, researchers, and clinicians (Heffernan et al. 2016).
- 2) Extant PROMs that were identified in two published systematic reviews (Seekins et al. 2012; Granberg et al. 2014).
- 3) The International Classification of Functioning, Disability, and Health (ICF) Core Sets for Hearing Loss (Danermark et al. 2013).

Subsequently, hearing-related participation restrictions were defined as the difficulties an individual with hearing loss experiences with authentic involvement in social situations. The term ‘authentic involvement’ was used because adults with hearing loss can appear to participate in social situations without being truly engaged, such as by pretending to follow a conversation (Heffernan et al. 2016). The term ‘social situations’ was used because the conceptualisation process demonstrated that hearing-related participation restrictions primarily occur in the social arena. Furthermore, ‘social situations’ is more precise and measurable than ‘life situations’.

A conceptual model of hearing-related participation restrictions was also developed, which contained three domains:

- 1) *Behaviour* - problems with performing actions in a social context due to hearing loss (e.g. difficulty with group discussions).
- 2) *Emotion* - negative feelings experienced in a social context due to hearing loss (e.g. feeling isolated at get-togethers).
- 3) *Identity* - negative social attributes perceived as stemming from hearing loss (e.g. being seen as unfriendly).

Each domain contained a range of subdomains (Please find this material at <http://tandfonline.com/doi/suppl>). Forty-nine items were generated to represent these subdomains, using the words and phrases of patients where possible. This included 26 behaviour items, 15 emotion items, and 11 identity items. The number of items associated with a domain was in proportion to the relevance of that domain to the target construct (Clark and Watson 1995). The behaviour items were positively-worded, whereas the emotion and identity items were negatively-worded (see Figure 1). The behaviour items were accompanied by an 11-point self-efficacy response scale, whilst the emotion and identity items were accompanied by an 11-point agree/disagree response scale (Rattray and Jones 2007; Sheer 2014). This first iteration of the questionnaire was entitled the SPaRQ-49.

[Figure 1 near here]

Once the SPaRQ-49 had been created, it was important to thoroughly evaluate its content in order to identify and rectify any flaws that could diminish its quality. Therefore, the first aim of this study was to evaluate the content of the SPaRQ-49 in terms of the following criteria:

- 1) *Relevance* - representative of hearing-related participation restrictions.

145 **2) Clarity** - easy to understand and interpreted as the questionnaire developers intended.

146 **3) Comprehensiveness** - captures all of the important aspects of hearing-related participation
147 restrictions.

148 **4) Acceptability** - inoffensive and not in any way intrusive to adults with hearing loss.

149 **5) Responsiveness** - sensitive to clinically relevant changes in hearing-related participation
150 restrictions.

151 The second aim was to improve the content of the SPaRQ-49 by making any necessary
152 amendments, such as introducing new items.

153 **Methods**

154 ***Design***

155 Content evaluation (i.e. pre-testing or content validation) is an essential component of PROM
156 development (Brod et al. 2009). It facilitates the assessment of two important measurement
157 properties: **(1)** content validity, or the relevance and comprehensiveness of the content of the
158 PROM and **(2)** respondent burden, or the degree to which the PROM poses a challenge for
159 respondents in terms of length, complexity, and literacy demands (Reeve et al. 2013).

160 Content evaluation typically involves key stakeholders appraising every element of a
161 questionnaire (e.g. items, response scale) against specific criteria (e.g. relevance, clarity)
162 (Haynes et al. 1995; Brod et al. 2009; Reeve et al. 2013). The PROM can then be amended
163 before it undergoes psychometric evaluation.

164 In this study, two prominent content evaluation techniques were used. Firstly, adults
165 with hearing loss (AHLs) participated in cognitive interviews. These are individual, semi-
166 structured interviews that uncover respondents' thought processes when completing a
167 questionnaire. For example, they reveal how respondents interpret the wording of the items or
168 how they decide which response category to select (Conrad and Blair 1996; Drennan 2003).

Cognitive interviews can be retrospective (i.e. conducted immediately after respondents have completed the questionnaire) or concurrent (i.e. conducted whilst respondents are completing the questionnaire) (Drennan 2003). Retrospective interviews were used because they examine whether respondents can follow the instructions and successfully complete a self-administered questionnaire (Willis 2004). Secondly, a panel of subject matter experts (SMEs), who had relevant clinical or academic qualifications and experience, completed a survey in which they evaluated the relevance, clarity, comprehensiveness, and responsiveness of the SPaRQ-49 (Haynes et al. 1995; Grant and Davis 1997; McGartland Rubio et al. 2003).

Participants

Adults with hearing loss

The inclusion criteria were self-reported: (1) hearing loss, (2) aged 18 years or older, (3) good written and spoken English language ability, and (4) normal or corrected-to-normal vision.

The exclusion criteria were self-reported: (1) cognitive decline or dementia that would necessitate assistance in completing a questionnaire and (2) profound hearing loss.

A convenience sampling strategy was used (Patton 1990). Potential participants were sought from the NIHR Nottingham Biomedical Research Centre (BRC) participant database. In total, 22 potential participants were contacted via post, of whom 14 participated in the study (see Table 1). Recruitment ceased when the research team determined that data saturation had been reached. This was the point at which no new themes or problems with the SPaRQ-49 were identified through an examination of field notes and preliminary data analysis (Leidy and Vernon 2008). The majority of AHLs had gradual-onset hearing loss and all owned hearing aids. Two individuals provided reasons for not participating, which were work commitments and health problems.

Subject matter experts

The inclusion criteria for the SMEs were identical to inclusion criteria 2-4 for the AHLs. A purposeful sampling strategy was used (Patton 1990; Grant and Davis 1997). Specifically, clinicians and academics who had expertise in adult aural rehabilitation and/or outcome measurement, as demonstrated by their academic qualifications, clinical qualifications, or publication history, were recruited from the professional network of the research team. It is recommended that SME panels have approximately 6-20 participants (McGartland Rubio et al. 2003). In this study, 29 potential participants were contacted via email, of whom 20 participated in the study (see Table 2).

Study procedure

This research was approved by the North East - Tyne and Wear South Research Ethics Committee, UK and Research and Innovation, Nottingham University Hospitals NHS Trust, UK.

Pilot testing

A pilot cognitive interview was conducted by the lead author (EH) with a NIHR Nottingham BRC Patient and Public Involvement (PPI) representative who had hearing loss. He made several valuable suggestions regarding study design. In particular, he advised that certain cognitive interview techniques (e.g. open-ended questions) would be less artificial and intrusive to AHLs than others (e.g. observation, thinking-aloud). Two NIHR Nottingham BRC researchers, who were not involved in the study, completed a pilot SME survey and suggested some minor alterations.

Cognitive interviews

Each participant attended the NIHR Nottingham BRC for their study session, which lasted approximately two hours. Written informed consent was obtained prior to the start of each session. The participants self-administered the SPaRQ-49, which took approximately 30

minutes. They were then interviewed by EH, who had formal training in and experience of interviewing, including interviewing AHLs (see Pearson et al. 2012; Heffernan et al. 2016). The interview schedule was flexible, yet its core content remained the same across each interview (Please find this material at <http://tandfonline.com/doi/suppl>). The interviews lasted 45 minutes on average and were audio-recorded and transcribed verbatim. The participants then completed a demographics questionnaire and the Davis et al. (2007) Strand 2 Screening Questionnaire for hearing loss. The participants were offered an honorarium of £10GBP and their travel expenses were reimbursed.

Subject matter expert survey

The SMEs completed an online survey, which took approximately one hour and 30 minutes. They answered a series of closed-ended and open-ended questions in which they evaluated the proposed factor structure, response scales, comprehensiveness, and responsiveness of the SPaRQ-49. They also rated the relevance and clarity of each SPaRQ-49 item using the following scale: 1= '*Does not fulfil criterion*', 2= '*Major revisions needed*', 3= '*Minor revisions needed*', 4= '*Fulfils criterion*' (Haynes et al. 1995; Grant and Davis 1997; McGartland Rubio et al. 2003). There were a mixture of optional and mandatory questions. Completion of the survey served as informed consent. The SMEs were offered an honorarium of £10GBP.

Data analysis

IBM SPSS Statistics for Windows Version 22.0 and QSR International's NVivo 10 Software were used to organise and analyse the data. Anonymised identification codes were assigned to each AHL (e.g. AHL1) and SME (e.g. SME1). For the SME survey data, descriptive statistics and frequencies were calculated for the closed-ended questions. The written comments from the open-text boxes were summarised and reported.

The cognitive interview data were analysed by EH using Braun and Clarke's (2006) thematic analysis procedure. The analysis was deductive, as the themes (i.e. relevance, clarity, comprehensiveness, and acceptability) were derived from the content evaluation literature, where it has been recommended that these criteria be examined (Brod et al. 2009; Mokkink et al. 2012). Deductive (i.e. theoretical) thematic analysis is a 'top-down' approach that is based on a pre-existing framework or the researcher's analytical interests. This contrasts with inductive thematic analysis, which is a 'bottom-up', data-driven process. The deductive approach was selected because it is suited to answering a specific research question, whereas the inductive approach is suited to exploring the data to develop further research questions (Braun and Clarke 2006).

Within the relevance, comprehensiveness, and acceptability themes, the data were coded inductively. Within the clarity theme, there were inductive and deductive codes. The deductive codes came from Conrad and Blair's (1996) taxonomy of problems encountered by questionnaire respondents. According to the taxonomy, there are three stages of responding to an item:

- 1) *Understanding* - deciding what information is being requested and recognising how this information should be provided.
- 2) *Performance* - producing the information needed to respond through mental operations (e.g. computation, evaluation).
- 3) *Response formatting* - mapping the information produced in the performance stage onto the response scale.

The taxonomy also lists several types of problems that can occur in each of the three response stages:

- 1) *Lexical* - problems with knowing the meanings of words.

- 265 **2) *Inclusion/exclusion*** - problems with deciding whether or not particular concepts are
266 within the scope of the item.
- 267 **3) *Logical*** - problems with negation, repetition, complementarity, contradictions, and
268 tautologies.
- 269 **4) *Computational*** - information processing problems that do not fall into one of the other
270 problem categories, such as complicated syntax or mental arithmetic.
- 271 **5) *Temporal*** - problems relating to the time period or frequencies specified in the questions.
272 This problem type was not applicable to the SPaRQ-49.

273 Therefore, in terms of coding, if a participant did not recognise a medical term used in an
274 item, this would be coded as a 'lexical-understanding problem'.

275 To enhance the rigour of this analysis, a peer assessment was completed (Yardley
276 2008). Specifically, EH and a second researcher, who was not otherwise involved in the
277 study, independently applied the taxonomy to seven interview extracts that had proven
278 challenging to code. For example, it was difficult to determine whether certain extracts
279 described a problem in the 'performance' or 'response formatting' stage. EH and the second
280 researcher then met to compare their coding. In the majority of cases, their coding matched.
281 Any discrepancies were discussed and an agreement was made regarding which codes should
282 be applied. In addition, the preliminary results were discussed with the research team to
283 ensure that the analysis was not limited to the viewpoint or preconceptions of EH.

284 ***Amendments***

285 Amendments were made to the questionnaire based on the results of the data analysis.
286 Specifically, aspects of the questionnaire that were identified as problematic by two or more
287 AHLs in the cognitive interviews were reviewed (Brod et al. 2009). In addition, aspects of the
288 questionnaire that received less than perfect ratings or comments in the SME survey were

reviewed. Subsequently, three PPI representatives who had hearing loss completed and provided feedback on the revised questionnaire. This process helped to ensure that the amendments were effective.

Results

Relevance

The cognitive interviews showed that the AHLs felt that the majority of behaviour items were representative of their experiences. A small number items about employment, volunteering, community activities, and interacting with a significant other were irrelevant to some AHLs because these situations did not arise in their daily lives, irrespective of their hearing loss. For example, AHL14 (man, aged 69) reported that the questions were representative of his hearing difficulties, with the exception of one that asked about participation in training courses: *'I can answer all those questions... because it's asking something I know about... I don't do any... courses... so I put zero on that [question], but the rest of them - it's very good... it's a very good questionnaire'*.

Some AHLs saw the emotion and identity items as being highly relevant. For example, AHL8 (woman, aged 62) said of the identity items:

They're very relevant because... that's a different level of your hearing loss, isn't it? A different effect that it has is about... how people perceive you... You do get treated as though you're not quite on the planet... at times... so I think that's a very relevant part of the questionnaire.

Contrastingly, some AHLs stated that the emotion and identity items were less relevant than the behaviour items. They explained that hearing loss did not lead them to feel particularly emotional or insecure due to their personality, particularly their self-confidence or sense of humour. AHL10 (man, aged 80) said:

I found it difficult to evaluate myself [in the emotion section] ... I don't mind being left out of a conversation... I just... sit through it and then move onto the next one... this comes back to your personality, doesn't it?... I don't feel stressed... I don't feel upset... It's just one of those things.

The results of the SME survey supported the interview findings, with the majority of the SPaRQ-49 items obtaining median and modal relevance ratings of 4 (Please find this material at <http://tandfonline.com/doi/suppl>). The mean relevance ratings for the individual items ranged from 3.35 to 4. The SMEs also evaluated the proposed factor structure of the SPaRQ-49. The majority (n=13) agreed that hearing-related participation restrictions consist of the domains of behaviour, emotion, and identity. However, some SMEs (n=7) disagreed. In the written comments, some SMEs reported that the behaviour dimension contained some items that represented activity limitations, rather than participation restrictions. Two SMEs stated that the identity domain was the least relevant. SME4 (Head of Adult Audiology Service) said:

[I] don't think identity is as significant as behaviour and emotion... people rarely report the impact of identity... [Patient] needs... mainly focus on behaviour but, with appropriate discussion, are often associated with [the] emotional dimension.

Clarity

The cognitive interviews showed that the AHLs found the majority of items to be easy to understand. Nevertheless, some clarity problems were uncovered through the application of Conrad and Blair's (1996) taxonomy to the data. The most substantial of these was a computational-response formatting problem. Specifically, the majority of AHLs struggled to switch from using the self-efficacy response scale accompanying the positively-worded behaviour items to using the agree/disagree response scale accompanying the negatively-

worded emotion and identity items. Some AHLs did not observe that the response scale had changed and assumed that the self-efficacy scale was present throughout the entire questionnaire. Other AHLs did observe that the response scale had changed but did not know how to adjust their responses accordingly. AHL8 (woman, aged 62) said: *'the marking changed... and that made me... stop and think... I couldn't really understand why it had... swapped over'*. Similarly, AHL2 (man, aged 78) said: *'When you swapped over... I think it probably did trip me up... I thought... couldn't you keep [the response scale] the same... all the way through the questionnaire?'* Consequently, the AHLs often selected a response that did not accurately represent their views, such as mistakenly selecting 'Completely agree' instead of 'Completely disagree'.

A second computational-response formatting problem was identified. Specifically, a small number of AHLs reported that they would have completed the questionnaire more quickly and easily if the 11-point response scale had fewer (i.e. 5 or 6) options. AHL12 (man, aged 64) said: *'Too many options. It's like having a big menu: you can't make a decision'*. In contrast, several AHLs stated that they had no difficulty with the 11-point response scales. Also, an examination of their responses showed that all of the AHLs used a range of response options, rather than using only the extreme ends or the middle of the scale.

Several AHLs experienced an inclusion/exclusion-performance problem, which was that they found it difficult to determine the scope of certain items, such as determining whether they referred to noisy environments, quiet environments, or both. However, other AHLs did not experience this problem. Instead they selected answers that represented their typical experience across both noisy and quiet environments. A final example of a clarity problem uncovered through the taxonomy was a logical-performance problem, whereby some AHLs perceived that certain items were repetitive, rather than truly distinct from one another. For example, AHL3 (woman, aged 62) felt that several emotion items were repetitive and

recommended merging the items that asked about isolation and loneliness. In contrast, AHL8 believed that the isolation and loneliness items were distinct: *'I think it's... really important... because... there's a difference... being isolated is almost like [being] on an island watching, whereas [being] lonely is a very personal... sadness or... aloneness... And I... have actually answered them differently'*.

The interview findings were supported by the SME survey. The majority of the SPaRQ-49 items had median and modal clarity ratings of 4. The mean clarity ratings for the individual items ranged from 2.9 to 4. In their written comments, some SMEs recommended providing more contextual information in certain items, such as clarifying whether they referred to noisy or quiet environments.

In terms of the response scales, the majority of SMEs (n=11) stated that the self-efficacy response scale did not need to be changed, though several (n=7) reported that change was required. In addition, the majority of SMEs (n=16) stated that the agree/disagree response scale did not need to be changed, though a small number (n=2) asserted that change was required. The written comments showed that one SME thought that there were too many response options, whilst two SMEs felt that the response options at the midpoint of the scales should be labelled, rather than unlabelled.

Comprehensiveness

The AHLs regarded the SPaRQ-49 as being highly comprehensive, as it assessed their main hearing-related difficulties. AHL4 (man, aged 77) said: *'basically you've got it all. I don't think you need to change anything... I really don't... I was quite surprised [by] how comprehensive it is'*. Nevertheless, the interviews uncovered some potentially important participation restrictions that were missing from the questionnaire, including reduced

385 independence, difficulties with participating in lengthy conversations, and friction with
386 communication partners.

387 In the SME survey, the majority of SMEs (n=13) agreed that the SPaRQ-49 was a
388 comprehensive measure. For instance, SME15 (Hearing researcher/audiologist) said: '*there*
389 *are some really important questions in here, which I doubt ever get asked in the... time*
390 *constraints of clinic*'. However, some SMEs were unsure (n=5) or disagreed (n=1) that the
391 SPaRQ-49 was comprehensive. Some recommended introducing open-ended questions that
392 would allow the respondents to personalise the questionnaire. Another recommended
393 ensuring that the participation component of the ICF Core Sets for Hearing Loss had been
394 fully captured.

395 ***Acceptability***

396 The majority of AHLs regarded the SPaRQ-49 as appropriate and inoffensive. AHL4 said:
397 '*there's nothing personal about it... it's actually very good... It's not intrusive... or anything*
398 *like that*'. However, two identity items, which referred to being treated as a nuisance and
399 being perceived as rude, were flagged as being potentially off-putting. Specifically, AHL11
400 (woman, aged 73) said: '*I don't really like... how that's phrased... it's very negative*'. In
401 addition, one participant, AHL8, reported that completing the questionnaire evoked
402 unexpected thoughts and emotions:

403 *It was very thought-provoking... it actually makes you think about how you feel about*
404 *not hearing, which... most people try and avoid... it actually made me quite sad...*
405 *because it... brings home just how much you miss out... but actually it's quite good*
406 *because... it made me reflect... I didn't feel like I've been... ripped asunder.*

407 Responsiveness

408 The SMEs were asked whether they agreed that the SPaRQ-49 would be a responsive PROM.
409 Seven agreed and three disagreed. The majority (n=9) selected ‘*Don’t know*’ in response to
410 this question. SME3 (Lecturer/Hearing therapist) wrote: ‘*[Its] a little hard to know... some*
411 *questionnaires are designed to be sensitive to change but turn out not to be! However, I do*
412 *think it asks about some of the things we would like to see change as a result of [an]*
413 *intervention*’. SME17 (Hearing researcher/audiologist) suggested that the identity might be
414 the least responsive domain: ‘*Tapping into identity is a brave and good idea. I wonder [how]*
415 *much of this would be expected to improve as a result of our current interventions*’. Two
416 SMEs warned that responsiveness is somewhat contingent on the timing of follow-up
417 assessments. As participation restrictions tend to change slowly over time, long-term follow-
418 up assessments may be necessary.

419 Amendments

420 Several amendments were made to the questionnaire in light of the above findings. The most
421 substantial amendment entailed revising the behaviour items so that they were negatively-
422 worded and accompanied by the agree/disagree scale. This revision addressed the difficulties
423 experienced by AHLs when switching from answering the behaviour items to answering the
424 emotion and identity items. The agree/disagree scale was selected to be the sole response
425 scale in the questionnaire because it received a higher rating in the SME survey than the self-
426 efficacy scale and because it was applicable to all of the items, whereas the self-efficacy scale
427 was applicable only to the behaviour items.

428 Several items were revised to improve their clarity. Specifically, some items were
429 altered to ensure that they were sufficiently distinct from one another. For example, an item
430 about watching live events (e.g. concert) was differentiated from an item about watching

television. Some items that substantially overlapped with one another were merged, such as two items about community activities and volunteering. Some items were adjusted so that they included a greater degree of contextual information, such as information about the acoustic environment. One item, which concerned being perceived as rude, was removed because it was found to be both off-putting and unclear. Finally, a small number of new items were created to enhance the comprehensiveness of the questionnaire, including items about managing responsibilities, participating in lengthy conversations, and getting along with others. The resultant iteration of the questionnaire contained 53 items (i.e. SPaRQ-53).

Discussion

This study aimed to evaluate and amend the content of the SPaRQ, a new hearing-specific PROM, in order to maximise its content validity and minimise any respondent burden. The results demonstrated that the majority of the SPaRQ content was relevant, clear, comprehensive, and acceptable. This likely reflects the benefits of having used a literature review and a previous qualitative study with key stakeholders to generate this content (Brod et al. 2009). Nevertheless, a number of potential problems were identified. For example, it was necessary to re-construct several items to improve their clarity. In addition, a small number of new items were created to enhance the comprehensiveness of the measure.

The most substantial problem identified was that most AHLs struggled to switch from answering positively-worded behaviour items, which were accompanied by a self-efficacy response scale, to answering negatively-worded emotion and identity items, which were accompanied by an agree-disagree response scale. Many questionnaires include reverse-worded items (i.e. items that are worded in the opposite direction to the other items) as a means of preventing response biases, especially inattention and acquiescence (Rattray and Jones 2007; van Sonderen et al. 2013). However, in this study, this approach did not

circumvent response biases but rather led to confusion and inaccurate responding. Other studies have similarly demonstrated that reverse-worded items fail to inhibit response biases and instead cause confusion, frustration, errors, and careless responding (Carlson et al. 2011; van Sonderen et al. 2013). Furthermore, reverse-worded items can detrimentally affect the psychometric properties of a questionnaire, particularly internal consistency and factorial validity (Woods 2006; Carlson et al. 2011). Consequently, the revised SPaRQ omitted reverse-worded items and contained a single response scale.

Another potential problem was that some participants felt that the 11-point response scales had too many response options. In particular, some AHLs felt that it would have been easier to select a response had there been five or six options. However, most participants did not object to the 11-point response scales and all of the AHLs were able to understand and use these scales. Therefore, it was concluded that, whilst some participants had a preference for five or six options, this did not mean that they had a problem with 11 options. Furthermore, the literature shows that response scales with a broader range of options are associated with greater responsiveness, reliability, and validity (Alwin 1997; Cummins and Gullone 2000; Weng, 2004; Leung 2011). There is also evidence to suggest that most respondents have a discriminative capacity greater than six points, which means that valuable data can be lost by adopting a response scale that is not sufficiently fine-grained (Cummins and Gullone 2000). Consequently, the 11-point response scale was retained.

It is not uncommon for content evaluation studies to uncover potential problems that ultimately do not lead to an amendment or that cannot be amended without creating additional problems. One previous study concluded that some of the potential faults identified within a physical activity questionnaire should not be amended because these amendments were associated with drawbacks as well as benefits (Andersen et al. 2010). Indeed, the purpose of content evaluation research is not the design of a 'perfect' questionnaire, but is

instead the facilitation of informed decisions about questionnaire design ‘trade-offs’ through uncovering the advantages and disadvantages of different formats (Beatty and Willis 2007).

Recommendations

This research has highlighted the importance of conducting a content evaluation study as part of developing a new PROM. This process can uncover serious problems, particularly irrelevant, unclear, or offensive content, which can reduce the amount and quality of data collected by the measure (McGartland Rubio et al. 2003; Brod et al. 2009). Despite these benefits, to date, just a small number of hearing-specific questionnaires have undergone a rigorous content evaluation (e.g. Smith et al. 2011). Therefore, it is recommended that new and existing hearing-specific questionnaires be evaluated to confirm that they have adequate content validity and minimal respondent burden and thus meet the standards required of high-quality PROMs (e.g. Terwee et al. 2007).

It is important to note that, whilst content evaluation studies can provide valuable data on relevance, clarity, comprehensiveness, and appropriateness, they may be less informative when it comes to responsiveness. In this study, many SMEs found it difficult to assess the responsiveness of the SPaRQ. The AHLs were not asked about responsiveness because they were unlikely to be familiar with this concept. Consequently, the responsiveness of the SPaRQ can only be assessed statistically in the later stages of its development (see Terwee et al. 2007). If the SPaRQ is found to have poor responsiveness at that stage, it may be necessary to re-develop and re-validate the measure. Future research should investigate techniques for maximising and assessing responsiveness in the early stages of developing a PROM. One strategy that has already been identified is the use of fine-grained response scales.

This research has demonstrated that it is not advisable to use reverse-scored items and/or multiple response scales in a questionnaire. Furthermore, caution should be exercised when administering a battery of questionnaires to participants, as the various response scales, instructions, and formats could cause confusion. In addition, researchers should consider the potential emotional or psychological impact of their questionnaires. In this study, one AHL found completing the SPaRQ to be an emotional, thought-provoking experience. She suggested making future respondents aware that they might have a similar experience. It is recommended that researchers take this into consideration when preparing the documents that will accompany the study questionnaires, such as by including the contact details of appropriate support services in the participant information sheet or questionnaire booklet.

Finally, this research found that some participants viewed the emotion and identity items to be highly relevant, whilst others felt that the opposite was true. Previous research has shown that hearing loss can have a considerable impact on emotions and identity (Barker et al. 2016; Vas et al. 2017). Research should examine whether variables such as personality traits and demographic factors influence this impact. The next stage of developing the SPaRQ involves assessing the effect of demographics (e.g. age, gender) on responses to the items.

Limitations

A limitation was that the AHLs were recruited through convenience sampling, which is one of the least rigorous sampling methods (Patton 1990). Consequently, AHLs with certain characteristics (e.g. non-ownership of hearing aids) were under-represented. Furthermore, information about the educational attainment of the AHLs was not collected. Fortunately, the interviews revealed that they had a wide range of education levels and occupations. An additional limitation is that the participants were aware that EH was involved in developing the SPaRQ. Although they were encouraged to identify problems with the questionnaire, it is

possible that some participants were not completely comfortable in providing negative feedback to EH. Another potential limitation was that the thematic analysis was deductive, rather than inductive, which may have increased the risk of overlooking important results that did not fit within the pre-existing themes or the researcher's analytical interests. However, even when utilising the inductive approach, it is difficult for researchers to completely suppress their pre-conceptions and analytical interests (Braun and Clarke 2006). Finally, the peer assessment in this study was somewhat restricted. Ideally, the interview data should have been fully analysed by at least two researchers who were formally trained in this type of analysis (Conrad and Blair 1996). Unfortunately, this was not feasible within the timeframe of this study.

Conclusion

This study utilised cognitive interviews with AHLs and a survey of SMEs to evaluate and revise the content of the SPaRQ: a new hearing-specific PROM. This process helped to ensure that the SPaRQ had sufficient content validity and minimal respondent burden, which are important measurement properties for any PROM (Reeve et al. 2013). To date, this approach to PROM development is rare in the field of hearing research. This study highlights the value of this approach means of eliminating serious flaws from a questionnaire and substantially improving its quality prior to its use in quantitative research. The next stage of this research is a quantitative study in which a modern psychometric analysis technique, namely Rasch analysis, is used to further evaluate and refine the SPaRQ. The ultimate aim of this research is to produce a high-quality measure of hearing-related participation restrictions that is suitable for use in both research and practice.

549 **Acknowledgements**

550 We would like to thank Sandra Smith, Alex Barker, the participants, and the PPI
551 representatives for their contribution to this research.

552

553

554 **Declaration of interest**

555 This paper presents independent research funded by the National Institute for Health
556 Research (NIHR). The views expressed are those of the authors and not necessarily those of
557 the NHS, the NIHR, or the Department of Health and Social Care in the UK. The authors
558 report no conflicts of interest.

559 **Table 1. Demographic information of the adults with hearing loss**

Gender	n
Male	7
Female	7
Age	Years
Mean	69.29
SD	9.07
Range	51-81
Estimated hearing loss duration	Years
Mean	14.57
SD	14.06
Range	3-51
Employment status	n
Retired	10
Employed	4

560

Table 2. Demographic information of the subject matter experts

Gender	n
Women	17
Men	3
Country of employment	n
UK	11
The Netherlands	3
USA	3
Denmark	2
Canada	1
Profession	n
Audiologist	10
Hearing researcher	5
Hearing therapist	3
Health psychologist	1
Engineer	1
Current occupation	n
Hearing researcher	13
Lecturer	3
Professor	3
Audiologist	3
Head of adult audiology service	1
Hearing therapist	1

Figure 1. Example of items from the SPaRQ-49

Behaviour item with self-efficacy scale	Despite my hearing problem, I can:	Certain I cannot do this										Certain I can do this
	Attend get-togethers with my family	0	1	2	3	4	5	6	7	8	9	10
Emotion item with agree/disagree scale	Because of my hearing problem:	Strongly disagree										Strongly agree
	I worry about going to social gatherings	0	1	2	3	4	5	6	7	8	9	10
Identity item with agree/disagree scale	Because of my hearing problem:	Strongly disagree										Strongly agree
	I think that I look like I am not interested in talking to people	0	1	2	3	4	5	6	7	8	9	10

Supplemental Material 1 - Example of Subdomains within the Hearing-Related**Participation Restrictions Conceptual Model**

Domain	Behaviour	Emotion	Identity
Example subdomains	Family relationships	Frustrated	Unsociable
	Interacting with strangers	Worried	Incapable
	Recreation and leisure	Embarrassed	Foolish
	Employment	Lonely	Unfriendly
	Community life	Isolated	Impolite
	Social relationships	Stressed	Apathetic

601 **Supplemental Material 2 - Cognitive Interview Schedule**

Initial Question	Potential follow-up	Component assessed
Can you tell me how you found filling out this questionnaire?	- Can you tell me more about that? - How did you find the length of the questionnaire? Did you find the questionnaire to be tiring to fill out?	- General
Can you tell me what that question means to you?	- Can you put that question in your own words? - How would you phrase that question? - What did that question bring to mind? - Was this question relevant to you?	- Clarity - Relevance
Can you tell me what that word/expression means to you?	- Can you put that expression in your own words? - How would you say it?	- Clarity
Can you tell me how you chose an answer to that question?	- Why did you choose that answer? - What were you thinking about when you were answering that question?	- Clarity
Were there any questions that you found difficult to understand?	- Can you tell me more about that? What made it difficult?	- Clarity
Were there any questions that you felt were not relevant to you?	- Can you tell me more about that? - Were there any questions that were especially relevant to you?	- Relevance
Were there any parts of the questionnaire that you found off-putting?	- Were there any questions that you didn't like? Were there any words or terms that you didn't like? - Were there any questions that you felt were too personal? - Were there any questions that you felt were offensive?	- Acceptability
Is the questionnaire missing any important questions?	- Are all of the main ways hearing loss affects people included in the questionnaire?	- Comprehensiveness
Can you tell me what you think about the instructions that come with the questionnaire?	- Were any of the instructions difficult to understand? - Do any of the instructions need to be changed? - Do any new instructions need to be added?	- Clarity
Can you tell me how you found using the rating scale?	- Did you have any problems using the rating scale? - What do the options on the rating scale mean to you? - What does the middle of the rating scale mean? - How did you decide what option to choose?	- Clarity - Appropriateness of the response scale
Can you tell me what you thought about the change in the rating scale?	- Can you tell me more about that? When did you realise it had changed? - Did that make the questionnaire more difficult for you?	- Clarity - Appropriateness of the response scale
Can you tell me what think of the layout of the questionnaire?	- Are there any changes that need to be made to the layout?	- Clarity
What do you think of the questionnaire overall?	- What was your overall reaction to the questionnaire? - What parts did you like and why? - What parts did you dislike and why?	- Relevance - Acceptability

602

603

604

605

606

607

608

609

610

611

612

613 **Supplemental Material 3 - Summary of the Subject Matter Expert's Relevance and**
 614 **Clarity Ratings for the SPaRQ-49 Items**

Relevance						
Domain	Mean	SD	Median	Mode	Min	Max
Behaviour items	3.67	0.60	4	4	1	4
Emotion items	3.84	0.47	4	4	1	4
Identity items	3.81	0.45	4	4	2	4
All SPaRQ-49 items	3.75	0.54	4	4	1	4

Clarity						
Domain	Mean	SD	Median	Mode	Min	Max
Behaviour items	3.51	0.63	4	4	1	4
Emotion items	3.82	0.44	4	4	2	4
Identity items	3.76	0.45	4	4	2	4
All SPaRQ-49 items	3.66	0.56	4	4	1	4

615

616

617

618

619

620

621

622

623

624

References

- Alwin, D. F. 1997. "Feeling thermometers versus 7-point scales: which are better?" *Socio Meth Res* 25: 318-340.
- Andersen, L. G., Groenvold, M., Jørgensen, T. and Aadahl, M. 2010. "Construct validity of a revised physical activity scale and testing by cognitive interviewing". *Scand J Public Health* 38: 707-714.
- Aronson, J. K. and Ferner, R. E. 2005. "Clarification of terminology in drug safety". *Drug saf* 28: 851-870.
- Barker, A. B., Leighton, P. and Ferguson, M. A. 2017. "Coping together with hearing loss: a qualitative meta-synthesis of the psychosocial experiences of people with hearing loss and their communication partners". *Int J Audiol* 56: 297-305.
- Beatty, P. C. and Willis, G. B. 2007. "Research synthesis: the practice of cognitive interviewing". *Public Opin Q* 71: 287-311.
- Boothroyd, A. 2007. "Adult aural rehabilitation: what is it and does it work?" *Trends Amplif* 11: 63-71.
- Boynton, P. M., Wood, G. W. and Greenhalgh, T. 2004. "Hands-on guide to questionnaire research: reaching beyond the white middle classes". *BMJ* 328: 1433-1436.
- Braun, V. and Clarke, V. 2006. "Using thematic analysis in psychology". *Qual Res Psychol* 3: 77-101.
- Brod, M., Tesler, L. E. and Christensen, T. L. 2009. "Qualitative research and content validity: developing best practices based on science and experience". *Qual Life Res* 18: 1263-1278.

- 646 Carlson, M., Wilcox, R., Chou, C.-P., Chang, M., Yang, F., Blanchard, J., Marterella, A., Kuo, A. and
647 Clark, F. 2011. "Psychometric properties of reverse-scored items on the CES-D in a sample of
648 ethnically diverse older adults". *Psychol Assess* 23: 558-562.
- 649 Clark, L. A. and Watson, D. 1995. "Constructing validity: basic issues in objective scale
650 development". *Psychol Assessment* 7: 309-319.
- 651 Collins, D. 2003. "Pretesting survey instruments: an overview of cognitive methods". *Qual Life Res*
652 12: 229-238.
- 653 Conrad, F. and Blair, J. "From impressions to data: increasing the objectivity of cognitive interviews".
654 *Proceedings of the Section on Survey Research Methods, Annual Meetings of the American*
655 *Statistical Association*, 1996. 1-10.
- 656 Cummins, R. A. and Gullone, E. "Why we should not use 5-point Likert scales: the case for
657 subjective quality of life measurement". *Proceedings, second international conference on*
658 *quality of life in cities*, 2000. 74-93.
- 659 Danermark, B., Granberg, S., Kramer, S. E., Selb, M. and Möller, C. 2013. "The creation of a
660 comprehensive and a brief core set for hearing loss using the international classification of
661 functioning, disability and health". *Am J Audiol* 22: 323-328.
- 662 Davis, A., Smith, P., Ferguson, M., Stephens, D. and Gianopoulos, I. 2007. "Acceptability, benefit
663 and costs of early screening for hearing disability: a study of potential screening tests and
664 models". *Health Technol Assess* 11: 1-294.
- 665 Dijkers, M. 2010. "Issues in the conceptualization and measurement of participation: an overview".
666 *Arch Phys Med Rehabil* 91: S5-S16.
- 667 Drennan, J. 2003. "Cognitive interviewing: verbal data in the design and pretesting of
668 questionnaires". *JAN* 42: 57-63.

- 669 Eyssen, I. C., Steultjens, M. P., Dekker, J. and Terwee, C. B. 2011. "A systematic review of
670 instruments assessing participation: challenges in defining participation". *Arch Phys Med*
671 *Rehabil* 92: 983-997.
- 672 Ferguson, M.A., Henshaw, H., Maidment, D. and Heffernan, E. In press. "Evidence-based
673 Interventions for Adult Aural rehabilitation: that was then, this is now". *Semin Hear*.
- 674 Ferguson, M. A., Kitterick, P. T., Chong, L. Y., Edmondson-jones, M., Barker, F. and Hoare, D. J.
675 2017. "Hearing aids for mild to moderate hearing loss in adults". *Cochrane Database Syst*
676 *Rev* 9.
- 677 Granberg, S., Dahlström, J., Möller, C., Kähäri, K. and Danermark, B. 2014. "The ICF core sets for
678 hearing loss – researcher perspective. Part I: systematic review of outcome measures
679 identified in audiological research". *Int J Audiol* 53: 65-76.
- 680 Grant, J. S. and Davis, L. L. 1997. "Selection and use of content experts for instrument development".
681 *Res Nurs Health* 20: 269-274.
- 682 Haynes, S. N., Richard, D. and Kubany, E. S. 1995. "Content validity in psychological assessment: A
683 functional approach to concepts and methods". *Psychol Assessment* 7: 238-247.
- 684 Heffernan, E., Coulson, N. S., Henshaw, H., Barry, J. G. and Ferguson, M. A. 2016. "Understanding
685 the psychosocial experiences of adults with mild-moderate hearing loss: An application of
686 Leventhal's self-regulatory model". *Int J Audiol* 55: S3-S12.
- 687 Heinemann, A. W., Tulskey, D., Dijkers, M., Brown, M., Magasi, S., Gordon, W. and Demark, H.
688 2010. "Issues in participation measurement in research and clinical applications". *Arch Phys*
689 *Med Rehabil* 91: S72-S76.
- 690 Leidy, N. K. and Vernon, M. 2008. "Perspectives on patient-reported outcomes".
691 *PharmacoEconomics* 26: 363-370.

- 692 Leung, S.-O. 2011. "A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-
693 point likert scales". *J Soc Serv Res*, 37: 412-421.
- 694 McGartland Rubio, D., Berg-Weger, M., Tebb, S. S., Lee, E. S. and Rauch, S. 2003. "Objectifying
695 content validity: Conducting a content validity study in social work research". *Soc Work Res*
696 27: 94-104.
- 697 Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M.
698 and De Vet, H. C. 2012. *The COSMIN checklist manual*, Amsterdam:
699 VU University Medical Center.
- 700 Patton, M. Q. 1990. *Qualitative evaluation and research methods*, New York: Sage Publications.
- 701 Pearson, J., Hammond, M., Heffernan, E. and Turner, T. 2012. "Careers and talents not to be wasted:
702 Skilled immigrants' journeys through psychological states en route to satisfying
703 employment". *Journal of Management Development* 31: 102-115.
- 704 Rattray, J. and Jones, M. C. 2007. "Essential elements of questionnaire design and development". *J*
705 *Clin Nurs* 16: 234-243.
- 706 Reeve, B. B., Wyrwich, K. W., Wu, A. W., Velikova, G., Terwee, C. B., Snyder, C. F., Schwartz, C.,
707 Revicki, D. A., Moinpour, C. M. and McLeod, L. D. 2013. "ISOQOL recommends minimum
708 standards for patient-reported outcome measures used in patient-centered outcomes and
709 comparative effectiveness research". *Qual Life Res* 22: 1889-1905.
- 710 Salter, K., Jutai, J. W., Teasell, R., Foley, N. C., Bitensky, J. and Bayley, M. 2005. "Issues for
711 selection of outcome measures in stroke rehabilitation: ICF participation". *Disabil Rehabil*
712 27: 507-528.
- 713 Seekins, T., Shunkamolah, W., Bertsche, M., Cowart, C., Summers, J. A., Reichard, A. and White, G.
714 2012. "A systematic scoping review of measures of participation in disability
715 and rehabilitation research: A preliminary report of findings". *Disabil Health J* 5: 224-232.

- 716 Sheer, V. C. 2014. "A meta-synthesis of health-related self-efficacy instrumentation: problems and
717 suggestions". *J Nurs Meas* 22: 77-93.
- 718 Smith, S. L., Pichora-Fuller, K. M., Watts, K. L. and La More, C. 2011. "Development of the listening
719 self-efficacy questionnaire (LSEQ)". *Int J Audiol* 50: 417-425.
- 720 Terwee, C. B., Bot, S. D., De Boer, M. R., Van Der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L.
721 M. and De Vet, H. C. 2007. "Quality criteria were proposed for measurement properties of
722 health status questionnaires". *J Clin Epidemiol* 60: 34-42.
- 723 Van Sonderen, E., Sanderman, R. and Coyne, J. C. 2013. "Ineffectiveness of reverse wording of
724 questionnaire items: let's learn from cows in the rain". *PLoS One* 8: e68967.
- 725 Vas, V., Akeroyd, M. A. and Hall, D. A. 2017. "A data-driven synthesis of research evidence for
726 domains of hearing loss, as reported by adults with hearing loss and their communication
727 partners". *Trends Hear* 21: 1-25.
- 728 Weng, L.-J. 2004. "Impact of the number of response categories and anchor labels on coefficient
729 alpha and test-retest reliability". *Educ Psychol Meas* 64: 956-972.
- 730 Whiteneck, G. and Dijkers, M. P. 2009. "Difficult to measure constructs: conceptual and
731 methodological issues concerning participation and environmental factors". *Arch Phys Med*
732 *Rehabil* 90: S22-S35.
- 733 World Health Organization. 2001. *International classification of functioning, disability and health:*
734 *ICF*. Geneva: World Health Organization.
- 735 Willis, G. B. 2004. *Cognitive Interviewing: A Tool for Improving Questionnaire Design: A Tool for*
736 *Improving Questionnaire Design*. London: Sage Publications.
- 737 Woods, C. M. 2006. "Careless responding to reverse-worded items: Implications for confirmatory
738 factor analysis". *J Psychopathol Behav Assess* 28: 189-194.

- 739 Yardley, L. 2008. Demonstrating validity in qualitative psychology. *In*: SMITH, J. A. (ed.)
- 740 *Qualitative psychology: A practical guide to research methods*. London: Sage Publications.