
Article:
Carlton, J. orcid.org/0000-0002-9373-7663, Peasgood, T., Mukuria, C. orcid.org/0000-0003-4318-1481 et al. (15 more authors) (2022) Generation, selection, and face validation of items for a new generic measure of quality of life : the EQ Health and Wellbeing. Value in Health, 25 (4). pp. 512-524. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2021.12.007

Abstract

Objectives

This paper aims to describe the generation and selection of items (Stage 2) and face validation (Stage 3) of a large international (multi-lingual) project to develop a new generic measure, the EQ Health and Wellbeing (EQ-HWB™), for use in economic evaluation across health, social care and public health to estimate Quality-Adjusted life Years.

Methods

Items from commonly used generic, carer, social care and mental health quality of life measures were mapped onto domains/sub-domains identified from a literature review. Potential terms and items were reviewed and refined to ensure coverage of the construct of the domains/sub-domain (Stage 2). Input on the potential item pool, response options, and recall period was sought from three key stakeholder groups. The pool of candidate items was tested in qualitative interviews with potential future users in an international face validation study (Stage 3).

Results

Stage 2 resulted in the generation of 687 items. Pre-determined selection criteria were applied by the research team resulting in 598 items being dropped, leaving 89 items that were reviewed by key stakeholder groups. Face validation (Stage 3) tested 97 draft items and 4 response scales. 47 items were retained, 14 were modified while three were added to the candidate pool of items. This resulted in a 64-item set.

Conclusion

This international multi-culture, multi-lingual study with a common methodology identified many items that performed well across all countries. These were taken to the psychometric testing along with modified and new items for the EQ-HWB.

Highlights

- Currently, few generic measures for economic evaluation exist. This study describes the process of the item generation and face validation stages from the E-QALY project.
- The face validation stage was carried out in six countries. Generally, participants favoured brief items. However, for some items, having examples and more information on the contexts could be helpful.
- This was an initial validation test of items that should be used in the EQ-HWB measure for economic evaluation of health and social care interventions.

Introduction

The development of new measures requires several stages to identify the relevant domains and items as well as further stages to test the validity of items in relevant populations. This includes the key assessments of content validity (how well items reflect the scope of what the questionnaire is trying to measure[1] and face validity (how appropriate, relevant and understandable items and their response options are[2,3]. There has been increasing demand for detailed accounts of the steps undertaken during these early stages of developing measures.[4] This paper aims to describe the generation and selection of items (Stage 2) and face validation (Stage 3) of a large international (multi-lingual) project to develop a new generic measure, the EQ Health and Wellbeing (EQ-HWB™), that can be used in economic evaluation across health, social care and public health to estimate Quality-Adjusted life Years (QALYs). Brazier et al[5] fully outline the rationale and the theoretical approach for the EQ-HWB. It is recognised that measuring health alone ignores that many conditions impact outcomes beyond health.[6] Such measures have limited ability in capturing outcomes in social care, nor do they take into account the impact of conditions upon informal carers. Use of a single measure will allow for comparison of interventions that impact individuals across sectors, and avoid risk of double counting. Having a common measure that is suitable for use across health, social care and public health will provide better evidence to help support cross-sector decision making.[7] The EQ-HWB has been developed for adults. Potential future work will explore the suitability of the measure for proxy reporting and/or child-user versions.

The project encompassed five stages outlined in Figure 1. This included (Stage 1) a literature review to identify potential domains, (Stage 2) item generation, (Stage 3) cognitive debriefing to test the face validity of potential items, and (Stage 4) psychometric analysis of a paper and online survey of potential items. After this stage, a broad consultation exercise identified items

to be included in a long version of the measure (25 items) and a shorter version (9 items) of the EQ-HWB measure. Stage 5 was the valuation phase, where selected items are valued by members of the public (to obtain utility weights for use in the estimation of Quality Adjusted Life Years or QALYs). More information on the overview of development of the measure and previous and subsequent stages are reported elsewhere.[5,8,9]

Methods

A large qualitative review was undertaken in Stage 1 that identified seven themes (feelings and emotions, cognition, activity, self-identity, relationships and social connections, 'coping, autonomy and control' and physical sensations) with 32 sub-themes as important domains and sub-domains of the quality of life of patients, social care users and informal carers.[8] A candidate pool of items was generated for the domains/sub-domains (Stage 2) and these were then tested with potential future users in an international face validation study (Stage 3).

[Insert FIGURE 1 here]

The focus for the overall project was on different populations of health, social care and informal carers with specific emphasise on using the new measure for economic evaluation. It therefore required specific considerations in the context of item generation and face validation to ensure that items were fit for purpose.[10,11] The criteria that an item was required to meet drew on existing published criteria[12,13] which was adapted following consultation with the steering and advisory groups of this project to meet the specific needs of the project in creating a generic health, social care and carer related QoL preference based measure.[11]

<u>Stage 2: Generation of candidate items</u>

Stage 2 drew from the qualitative literature review themes and sub-themes in Stage 1.[8] There were four steps: a) sourcing items to map to the 32 sub-domains (seven domains); b) refinement and modification of items; c) review of items from stakeholder, advisory and Patient and Public Involvement and Engagement (PPIE) groups; and d) further refinement of items and response options.

Step 2a: Sourcing items to map to domains/sub-domains

Concepts and terms from the literature review, categorized in domains and sub-domains, were summarised and possible items were identified from existing questionnaires and item banks. Items from commonly used generic, carer, social care and mental health QoL measures were mapped onto the domains/sub-domains. Information on the source, relevant sub-domain(s), original item wording, alternative wording, response options and notes on whether there were potential problems with the item based on the criteria, such as covering more than one concept, were documented.

Step 2b: refinement and modification of items

Potential terms and items were reviewed by the research team to ensure coverage of the construct of the domains/sub-domain. Due to the potentially vast number of existing published items on health and QoL, application of the selection criteria began at early screening stages of item generation. Alternative wording was used to modify items (based on team discussions and consensus) where the original item did not fit the proposed structure or criteria for item selection of the new measure.

Step 2c: review of items from stakeholder, advisory and PPIE groups

Input on the potential item pool, response options, and recall period was sought from three key

stakeholder groups. The project PPIE group participated in a focus group session where they were asked to share their thoughts on each item. A second focus group was held with members of the National Institute for Health and Care Excellence (NICE) Citizen's Council who are members of the public including patients and social care users. Two researchers with experience in focus group methods facilitated the focus groups. The project international advisory group (consisting of industry, academics and developers of measures) also provided comments on the proposed item pool via an online survey. In the survey, background information was provided via a video and report, before participants were asked to highlight problematic items with reasons and to provide alternatives. The potential pool of items was also presented to NICE staff who were asked to provide feedback.

Step 2d: refinement of items and response options

Findings from step 2c were summarised in a spreadsheet and used to refine item wording (where appropriate) and reduce the number of items within the item pool to take forward into Stage 3. This included changing any ambiguous words, adding explanations and dropping any items that were considered particularly problematic based on the feedback received.

Stage 3: Face validation

Data collection

Face validation studies were conducted in six countries, Argentina, Australia, China, Germany, United Kingdom (UK), and United States of America (US). Semi-structured one-to-one cognitive interviews were undertaken with members of the public and carers, patients and social care users.[14] Participants were asked how they would interpret each question, their ability to respond to it as well as their preferences over similar questions with different framing or wording. They were also asked for alternative wording if they highlighted problems with the

proposed wording. Each participant saw only a subset of the domains with an overall total of 30-50 items. Items were shown in a questionnaire format (Figure 2). In some cases, different response options could apply i.e. frequency, severity, difficulty or agree-disagree and respondents were asked if they had a preference. All interviewers were provided with training documents and videos and a topic guide (see Supplementary Material). Primary investigators in each country were responsible for ensuring that interviews were undertaken in line with the protocol to ensure a level of consistency internationally. Interviews were conducted in the native language of the participant. A detailed summary of the findings were shared to the wider research team in English. Written informed consent was taken at the start of each interview. Participants completed a short survey (age, gender, ethnicity, any health condition they suffer from, any caring role they have, and EQ-5D-5L), though these questions were not compulsory. At the end of the interview, participants were compensated. All interviews were audio-recorded using an encrypted device and researchers also made brief notes. Ethical approval was obtained from the Institutional Review Boards and relevant Ethics Committees.

Three countries (Argentina, China and Germany) needed translation from English to the respective languages prior to face validity work. A single translation company undertook the translation following best practice guidelines with forward and back translation by different translators followed by input from the country research team alongside support from the UK team to ensure that the appropriate translations were used (i.e. steps 1-6 and 9-10 of the current best practice guidance[15]).[16] Topic guides were translated by the country teams.

[Insert FIGURE 2 here]

Participant sample

Patients, social care users, carers (both formal and informal) and members of the general population were invited through different channels in every country (Table 1).

Data analysis

Data generated from the interviews were analysed systematically by considering and documenting all feedback/comments reported by the respondents. Data was recorded on a piloted extraction sheet (see Supplementary Material, Table 1) where item meaning, comprehensibility, item preference, response option preference and suggested alternatives were recorded. Although interviews were not transcribed verbatim, analysis involved listening to interview recordings and revising notes to ensure immersion in the qualitative data. The researcher that conducted the interview made notes for each item related to the meaning/interpretation of the item, any positive or negative points raised, any suggested alternatives and preferred items/response options where this was applicable. This information was combined to provide information on the items, including which items to drop (and therefore not be tested in Stage 4), take forward (with or without refinement) to Stage 4, and suitability of response options. Each country independently rated each of the items, and provided recommendations about which items to retain. The results were then summarised across countries. Self-reported characteristics were used to assess whether particular issues with items arose more in certain groups than others.

Results

Stage 2: Generation of candidate items

Step 2a: mapping of items to domains/sub-domains

After reviewing a large pool of items (n=2197) against the selection criteria, a total of 687 items were collated. Of these, 458 items were extracted from the generic preference and non-preference-based measures in health and social care as well as wellbeing measures while 229 were drawn from item banks and other measures (Supplementary Material, Table 2). Some concepts such as 'support', 'stigma' and 'cognition' were identified as being inadequately covered at this stage. Targeted measures and a recent study reviewing measures for assessing wellbeing, happiness and QoL were used to help identify more items to address these gaps.[17]

Step 2b: refinement and modification of items

A more detailed review of the items by the team against the selection criteria resulted in many of the items (n=598) being dropped from further consideration. There were a number of reasons for dropping items. Many of the items were similar in nature covering the same concepts e.g. different ways of asking about pain and those that were considered to be suitable for a measure that would be used in valuation were selected. There were also items that asked about two aspects e.g. impact of pain on functioning that we sought to avoid. In the initial draft item selection, both positively and negatively phrased items were included with further consideration on this issue undertaken in later stages of the project. There was overlap between items related to different sub-domains within and across domains. Social engagement items were related to items in other relationship and activity items; autonomy items were related to control and activity items; thinking clearly was related to other cognition items – therefore these sub-domains were not explicitly taken forward. Items identified for the self-worth/respect sub-domain were split into confidence and self-worth sub-domains.

A number of aspects were taken into consideration around the choice of response options. This included whether or not frequency or intensity best distinguished the level of attainment for a

sub-domain and the specific wording used. The number of levels were considered based on existing measures, evidence from the literature and judgement within the research team; a default position of five levels was adopted.

Recall periods adopted for self-reported measures vary from today (or yesterday) to last month. The recall period can impact on applicability, which may cause missing items (resulting in missing data).[18] Very short recall periods such as today/yesterday may mean that respondents are not experiencing the issues raised on the particular day.[12,19] Additionally, capturing broader QoL domains such as coping, control and loneliness may require a slightly longer recall period. As noted by Norquist et al[19] 'Longer recall periods may be necessary…when consideration, and integration of events over some period of time is required to reasonably report on the underlying patient reported outcome (PRO) concept (e.g., social functioning)'. On the other hand, respondents may not remember information accurately over a long recall period and will only report the most salient information rather than 'on average'.[12] The need to generate a measure that could be used to track progress following acute events (such as stroke or fracture) in which QoL may change fairly rapidly, also makes longer periods of time problematic. A default position of seven days was adopted at the outset, with regular consideration as to whether this would be most suitable for each item.

Step 2c and 2d: review and refinement of items

The results from the face-to-face focus group sessions with NICE Citizens Council (n=5) and the PPIE group (n=7) were combined with the responses from the online survey of advisory group members (n=28 responses received). Feedback from the consultation frequently focused on adherence/consistency of application of the selection criteria although there was feedback on specific items. Participants provided views on the different items including

interpretation and value of including the questions. The advisory group noted 'I felt' as more subjective than 'I was', which may also be considered for some items as a clinical diagnosis. The item 'I felt cross' was considered problematic by the PPIE and 11 members of the Advisory Group and hence dropped. Items from the Adult Social Care Outcomes Toolkit (ASCOT) were identified as problematic for generic use as they were tailored towards recipients of care. The sub-domains around guilt/shame and burden were dropped during early consultation due to social desirability concerns. Further detail of the results of the PPIE results are shown in the Supplementary Material. Ninety-seven items were taken forward into face validation.

Stage 3: Face validation

Face validation studies were conducted between April 2018 and February 2019. Table 1 shows participant characteristics for the face validity study for each of the participating countries. A total of 170 interviews were conducted with patients (n=79), social care users (n=23), carers (both formal and informal, n=50) and members of the general population (n=18).

[INSERT TABLE 1 HERE]

A summary of the common and core findings for each of the seven domains are outlined below and summarised in Table 2.

Domain specific-findings

Of the 97 draft items taken into Stage 3, 36 items were eliminated based on the evidence in this stage. Three additional items were added. This resulted in a 64-item set (see Figure 3).

[INSERT FIGURE 3 HERE]

Activity

This domain aimed to capture functioning and covered self-care, enjoyable or meaningful activities/roles, mobility, communication (speech), hearing and vision. Twenty-four potential items were tested and eleven were dropped, while one was added (Figure 3). Questions which referred to what individuals 'wanted' to do versus 'needed' were interpreted correctly with the former referring to what was preferred and the latter to activities that were essential such as activities of daily living. However, for some items, there was ambiguity due to differences in interpretation, brevity and the lack of context. For example, some items were interpreted in different ways to what was intended e.g. 'communicate' inferred to mean methods of communication – telephone, conversation, text and email; skill in getting a message across effectively; the response of others (e.g. clinical staff not listening to them). This does not link to the original construct of hearing and speaking and points to ambiguity as to what respondents' answers would be referring to. The term 'self-care' was not commonly used to mean things like washing/dressing. In mental health, self-care was interpreted to mean the things that they did to improve their wellbeing, rather than in terms of physical self-care (i.e. washing, dressing). Similarly, self-care was seen as arising from both physical limitations and resource limitations (e.g. lack of time).

Including aspects of 'receiving help' was problematic even in groups where help could have been received (i.e. patients) therefore this was rephrased. The items aimed to distinguish between personal care outcomes attained over the last week (what actually happened) and the respondent's ability to attain personal care outcomes independently (what they would have

achieved if they didn't have care/support). These items created ambiguity in interpretation from respondents who didn't receive any care, or would have benefited from additional care/support.

The relevance of some items was also highlighted. This included comments around what could be reasonably expected e.g. 'everyone experiences boredom' or 'unrealistic to expect people to be able to do what they want'. There were also issues with questions related to self-care and receiving help for some carers who did not know why they would be asked these questions.

Autonomy

This domain covered coping and control and was mainly testing different ways of asking the same question. Seven potential items were tested, two were dropped and one new item was added (Figure 3). There was a preference for items that had more information e.g. coping with day-to-day life rather than just coping. An item which provided a definition of control was found to be helpful by many of the respondents.

Cognition

Concentration, memory and confusion were covered in this domain. Seven potential items were tested and two were dropped (Figure 3). Most participants understood the questions and said they would be able to answer them. 'Memory' was considered to be a long-term issue and not something in the context of 7 days. Some respondents interpreted this to be referring to dementia with some questioning whether this would be something that could be answered i.e. 'would I know that I have memory loss'.

Feelings and Emotions

This domain covered sadness, happiness, worry, hope and hopelessness, anger and frustration, vulnerability and safety, and guilt/shame. Twenty-five potential items were tested and nine were dropped (Figure 3). Many of the items were interpreted correctly and respondents could answer them, though there were issues with some. In the happiness/depression sub-domain, some respondents felt that the top end of 'happy' and 'enjoyed life' were unrealistic i.e. 'no one enjoys life all the time'. The term 'depressed' was interpreted to mean having a clinical diagnosis by some respondents. In the hope/hopelessness sub-domain, the item on 'life not worth living' was considered quite negative. Looking forward to each day was not considered to be something that individuals did every single day, while 'look forward to' needed further information in some countries. 'Safe' and 'secure' were considered to be ambiguous terms in the safety sub-domain while 'relaxed' was considered to be a physical state in the anxiety/calm domain.

Physical Sensations

This domain covered pain, discomfort, sleep problems and fatigue. Eight potential items were tested, most of which performed well in face validity and only one was dropped (Figure 3). Discomfort was often interpreted to include mild pain. The term 'physical' was added to pain and discomfort items to distinguish this from mental health-related aspects.

Relationships

This domain covered loneliness, social engagement, stigma, support, positive relationships and relationships, belonging and connectedness, and burden to others. Sixteen potential items were tested, with many performing well in terms of interpretation and ability to respond to them and only five were dropped (Figure 3). Social support framed as 'support' or 'by other people' resulted in some ambiguity. 'Support' was unclear while 'other' resulted in respondents

considering people who were not those they saw regularly. ‘Disagreements and conflict’ was considered problematic as it focused on two issues and had mixed interpretation in terms of impact on QoL as some respondents thought of it as a positive to be able to have disagreements (UK and Australia). ‘Got on’ was colloquial and did not translate well. The term ‘judged’ was also ambiguous and not necessarily negative in all interpretations.

Self-identity

This domain aimed to cover feelings of confidence and self-worth, and being treated with dignity/respect. Ten items were tested and six were dropped including one sub-domain, dignity/respect (Figure 3). ‘Confidence’ had broad interpretations some of which were relevant. However, many of the other items in this domain were problematic. Dignity was linked to respondents' own behaviour rather than the behaviour of others while respect was linked to manners or very specific incidents. Therefore, this sub-domain was dropped. ‘Feeling valued/useful’ was not relevant to older people due to how the terms were interpreted i.e. doing tasks or being paid. ‘Feeling good’ had some irrelevant interpretations e.g. ‘how I look’ while others were related to physical health i.e. ‘I felt well’.

Common Findings

Respondents found it useful to have examples of the construct being measured – and this was a common finding across the different domains. Brief items could be answered but respondents wanted information on context and this was true across different countries. There was also a preference for simpler layouts in presenting questions.

Although there were some differences in response option preferences e.g. frequency over severity, this was often mixed and respondents were often unable to say why they preferred

one option to another. Recall periods were sometimes considered too short for particular constructs such as coping, control or irrelevant such as hearing where the loss is permanent. Completion instructions for the draft measure, including the recall period, were usually displayed at the top of the page or table. These were often ignored or forgotten by participants.

Combining the evidence to inform the content of the psychometric survey (Stage 4)

The results of Stages 2 and 3 were used to inform the selection of items taken forward to Stage 4 (psychometric survey)[9] (Table 2).

Discussion

This project aimed to develop a broader generic measure of QoL for use in economic evaluation that would be relevant for use across health and social care. Methods of development drew upon current good practice for measure development, covering multi-country, multi-lingual and multi-cultural considerations.[4,13,15,20] The generation of items based on terms from the qualitative review[8] and items from existing health and wellbeing measures resulted in 687 candidate pool of items from a list of 2,197 potential items. Items were identified for 28 sub-domains across seven domains. This approach allowed for full consideration of the relevance, comprehensiveness and comprehensibility (i.e. content validity) of the new measure.

Stage 3 incorporated an ambitious multi-country face validation exercise to further test and examine the suitability of the proposed item pool and response options. 97 items were tested in the face validation and 47 items were retained, 14 were modified while three were added to the candidate pool of items for consideration in further stages. One sub-domain was dropped. The approach benefited early in the development phase of the measure from a multi-culture,

multi-lingual approach with common methodology employed across different countries, which was important in considering wider audiences who may use the measure.

The results were used to help inform the reduction of the item pool to take forward to Stage 4 (psychometric survey) and were used as evidence to inform final item selection for the EQ-HWB measure. Many items were identified as being potentially problematic in face validity interviews across the different groups. Short items without additional context raised concerns and uncertainties about their scope yet longer items risked problems with readability. Using different population groups was important as some items worked better in some groups compared to others. For example, being able to communicate well, from a patient perspective has a physical emphasis, for some non-patients/carers this is interpreted as how successfully they reveal communication skills.

The project was not without its challenges. Logistical difficulties associated with ethical and governance approval processes across the included countries made iterative decision-making challenging. Whilst general population, patient, carer and social care perspectives were sought across the whole project, this was not achieved for all countries. Recruitment from social care was completed in three of the six countries (Argentina, England and Germany). The steps undertaken in the development of potential items and response options was robust and followed recognised best practice. This study did not undertake a qualitative study to generate items as advocated in the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN).[21] Instead, data from existing evidence (including published qualitative reviews and established measures of health and wellbeing) was used which had the advantage of drawing from a broader range of voices including different mental and physical health patient groups, carers of different types of individuals and users of social care. Audio-

recordings from discussions with PPIE or stakeholder groups and face validation studies were not transcribed verbatim as recommended in COSMIN.[21] Whilst verbatim transcription was not undertaken, audio recordings were used to complete data extraction from the interviews themselves. Given the tight focus of the interviews on cognitive debriefing of pre-determined items transcription was not considered necessary. Resource and time implications were considered, however the primary reason was one of minimising research waste and the ethical implications of undertaking research with no clear rationale. It was viewed to be more important to check interpretation across a broad sample.

Conclusion

A candidate pool of items was identified and selected for testing in face validation across six countries to cover a broad range of content important to patients, social care users and informal carers around the world. In these initial stages we exhaustively searched items, mapped them to domains and sub-domains, and carried forward a successful face validation of an initial item pool. Though there were some discrepancies among six countries there were useful common findings to select items for the next stage. In doing this, items were identified that were considered appropriate and understandable across all included groups of participants and across different countries and cultural contexts. The international evidence was used to support decision-making for item retention and elimination for subsequent stages of the EQ-HWB development. The EQ-HWB has a potential for becoming a valuable addition to the supply of QoL measures in research and economic evaluation across health, social care and public health around the world.

References

1. McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. Second edition. Oxford University Press, New York. 1996.
2. Connell J, Carlton J, Grundy A, Taylor Buck E, Keetharuth AD, Ricketts T, Barkham M, Robotham D, Rose D, Brazier J. The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). Qual Life Res. 2018;27(7):1893-1902.
3. Johnson RL, Morgan GB. Survey scales: A guide to development, analysis, and reporting. Guilford Publications. 2016
4. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for industry: patient-reported outcomes measures: use in medical product development to support labeling claims. Issued December 2009. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM19328.pdf. Accessed 19th May 2021.
5. Brazier JE, Peasgood T, Mukuria C, et al. "Development of a new generic measure of health and wellbeing for estimating Quality Adjusted Life Years: the EQ Health Wellbeing (EQ-HWB)" Value in Health Submitted in themed issue
6. Mitchell PM, Al-Janabi H, Richardson J, Iezzi A, Coast J. The relative impacts of disease on health status and capability wellbeing: A multi-country study. PLoS One. 2015; 10(12): e0143590. doi: 10.1371/journal.pone.0143590
7. Brazier JE, Rowen D, Lloyd A, Karimi M. Future directions in valuing benefits for estimating QALYs: Is time up for the EQ-5D? Value Health. 2019;22(1):62-68. doi: 10.1016/j.jval.2018.12.001
8. Mukuria C, Connell J, Carlton J, Peasgood T, Scope A, Clowes M, Rand S, Jones K, Brazier JE "A targeted review of qualitative evidence on domains of quality of life important for

patients, social care users and informal carers to inform the development of the EQ Health and Wellbeing (EQ-HWB)" Value in Health Submitted in themed issue

9. Peasgood T, Mukuria C, Brazier JE, Marten O, Kreimeier S, Luo N, Mulhern B, Pickard AS, Augustovski F, Greiner W, Engel L, Yang Z, Gibbons L, Monteiro A, Kuharic M, Belizan M, Bjørner J "Developing a new generic health and wellbeing measure: psychometric survey results for the EQ Health and Wellbeing (EQ-HWB)" Value in Health Submitted for themed issue
10. Peasgood T, Mukuria C, Carlton J, Connell J, Devlin N, Jones K, Lovett R, Naidoo B, Rand S, Rejon-Parrilla JC, Rowen D, Tsuchiya A, Brazier J. What is the best approach to adopt for identifying the domains for a new measure of health, social care and carer-related quality of life to measure quality-adjusted life years? Application to the development of the EQ-HWB? Eur J Health Econ. 2021 Apr 28. doi: 10.1007/s10198-021-01306-z.
11. Peasgood T, Mukuria C, Carlton J, Connell J, Brazier J. Criteria for item selection for a preference-based measure for use in economic evaluation. Qual Life Res 2021;30(5):1425-1432. doi: 10.1007/s11136-020-02718-9.
12. Bradburn NM, Sudman S, Wansink B. Asking questions: the definitive guide to questionnaire design--for market research, political polls, and social and health questionnaires. John Wiley & Sons; 2004 May 17.
13. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 2014. 5th edition.
14. Tourangeau R. Cognitive science and survey methods: a cognitive perspective. In: Jabine T, Straf M, Tanur J, Tourangeau R, editors. Cognitive Aspects of Survey Design: Building a Bridge between Disciplines. Washington, DC: National Academy Press; 1984. pp. 73–100.

15. Arafat S, Chowdhury H, Qusar HM, Hafez MA. Cross cultural adaptation and psychometric validation of research instruments: a methodological review. J Behav Health 2016;5(3): 129-136. DOI: 10.5455/jbh.20160615121755
16. Linton MJ, Dieppe P, Medina-Lara A. Review of 99 self-report measures for assessing well-being in adults: exploring dimensions of well-being and developments over time. BMJ Open. 2016;6(7):e010641.
17. Morris, Devlin, Parkin, Spencer (2012) Economic Analysis in Healthcare, 2nd Edition; ISBN 978-1-119-95149-0
18. Norquist JM, Girman C, Fehnel S, DeMuro-Mercon C, Santanello N. Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. Qual Life Res. 2012;21(6):1013-20. doi: 10.1007/s11136-011-0003-8.
19. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P. Principles of good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8(2):94-104.
20. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, Ring L. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. Value Health. 2011;14(8):978-88. doi: 10.1016/j.jval.2011.06.013.
21. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. Qual Life Res. 2021. doi: 10.1007/s11136-021-02822-4. Epub ahead of print.
22. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5 D-5 L value set for England. Health economics. 2018;27(1):7-22.

23. Ludwig K, Graf von der Schulenburg JM, Greiner W. German Value Set for the EQ-5D-5L. Pharmacoeconomics. 2018;36(6):663-674. DOI: 10.1007/s40273-018-0615-8.

24. Pickard AS, Law EH, Jiang R, Pullenayegum E, Shaw JW, Xie F, Oppe M, Boye KS, Chapman RH, Gong CL, Balch A, Busschbach JJV. United States Valuation of EQ-5D-5L Health States Using an International Protocol. Value Health. 2019;22(8):931-941.

Figure 1 Overview of the development of the EQ Health and Wellbeing (EQ-HWB™)

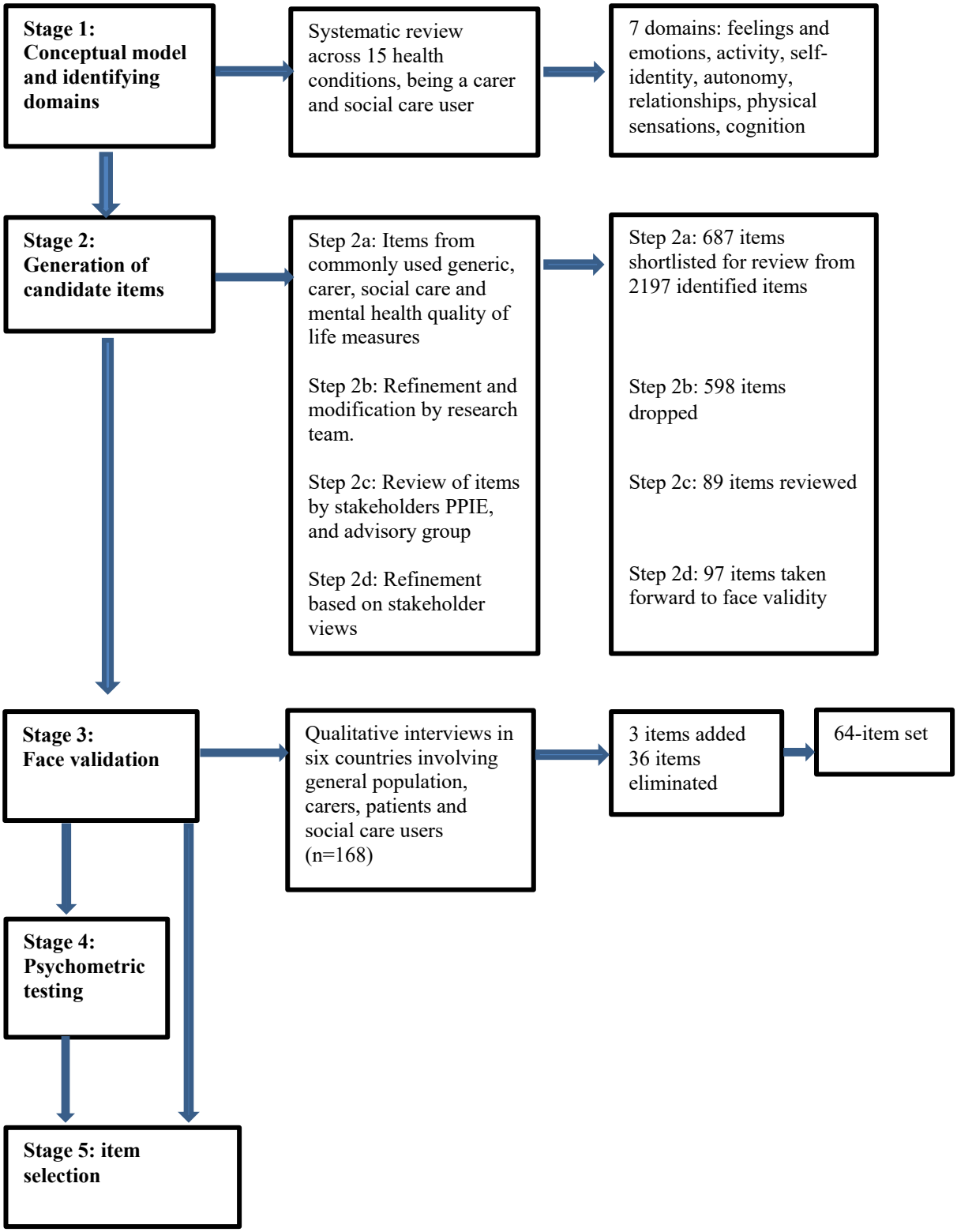

Figure 2 Example of display of items with potential response options

For each of the following statements, please tick one box that best describes your thoughts, feelings and activities over the **last 7 days**

	None of the time	Only occasionally	Some of the time / Sometimes	Often	Most or all of the time
1 I found it hard to concentrate	☐	☐	☐	☐	☐
2 I found it hard to focus my thoughts	☐	☐	☐	☐	☐
3 I found it hard to pay attention	☐	☐	☐	☐	☐
4 I had trouble thinking clearly	☐	☐	☐	☐	☐
5 I had trouble remembering	☐	☐	☐	☐	☐
6 I had trouble with my memory	☐	☐	☐	☐	☐
7 I felt/was confused	☐	☐	☐	☐	☐

Alternative Response Sets

Frequency: None of the time, Only occasionally, Some of the time / Sometimes, Often, Most or all of the time

Severity: Not at all, A little bit, Somewhat, Quite a bit, Very much

Level of Difficulty: No difficulty, Slight difficulty, Some difficulty, A lot of difficulty, Unable

Agreement: Strongly agree, Agree, Neither agree or disagree, Disagree, Strongly disagree

Figure 3 Summary of item modification following face validation

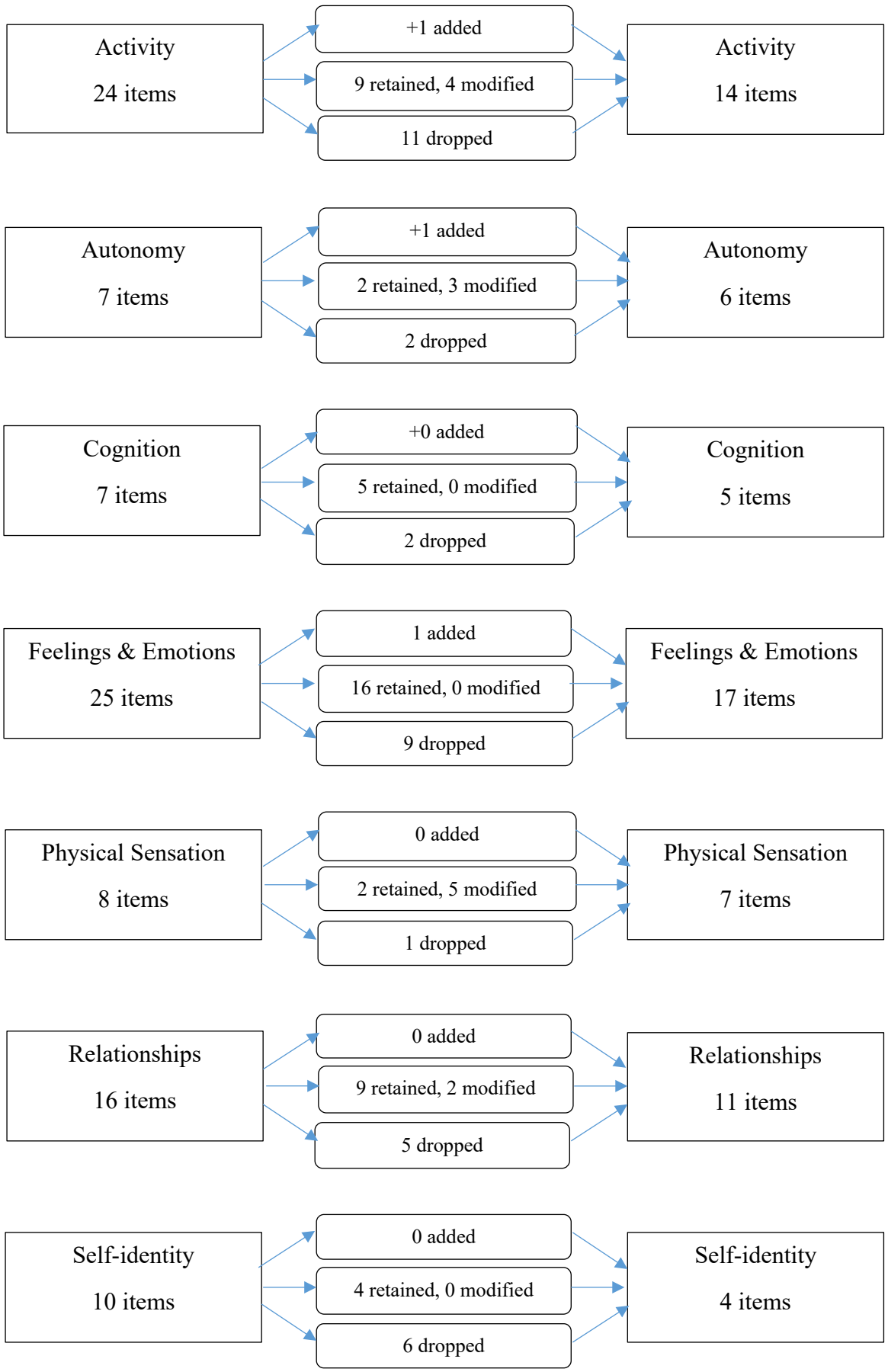

Table 1 Face validation participant demographics

	General public	Carers	Patients	Social-care users	Total	Age range Mean (SD)	Female (%)	EQ-5D country tariff utility value	EQ-VAS Mean (SD)
Australia	4	4	17	0	25	28-70 53.7 (14.1)	56	0.848 (0.131)†	N/C
	Participants were recruited through an external recruitment company (Stable Research). Purposive sampling was used to include individuals with various physical and mental health conditions, carers and members of the general public.								
Argentina	8	8	0	8	24	24-91 54 (20)	63	N/C	N/C
	Participants were recruited using different strategies. Known individuals were contacted (through local researchers' informal networks). A snowball sampling approach was adopted asking participants to help researchers to identify further individuals, and particularly social care users. Finally, we visited health promotion public facilities in the city of Buenos Aires ("Estaciones Saludables") to recruit users of those services.								
China	0	13	17	0	30	18-71	60	N/C	N/C

						37.73 (15.55)			
	Participants were recruited using a convenience sampling approach from two hospitals in Shanghai, No.10 Hospital of Shanghai and Zhongshan Hospital of Fudan University. Most participants were recruited from the outpatient services, some were recruited from inpatient services.								
England	6	13	18	8	45	23-95 60.4 (20.2)	58	0.78 (0.23)†	N/C
	Participants with physical health conditions were recruited from Sheffield Teaching Hospital Patient panels (Cardiovascular Patient Panel, Diabetes & Endocrinology Panel, Therapeutics & Palliative Care Panel, Online Public Advisory Panel, Motor Neurone Disease Panel, Stroke Panel). Mental health service users were recruited through RDaSH targeting mental health service users including those receiving drug and alcohol rehabilitation. Social care users were recruited through a day centre and residential care home (via Doncaster City Council). Carers were recruited through Sheffield Carers Centre via an email to their list and an advert on their website. Members of the general public were recruited through University of Sheffield volunteers list for staff but excluding academic staff and the School of Health and Related Research (where the research was conducted).								
Germany	0	12	8	7	27	21-30 yrs n=6 31-40 yrs n=6	70	0.85 (0.20)§	73.50 (19.68)

						41-50 yrs n=4 51-60 yrs n=7 61-70 yrs n=2 71-80 yrs n=2			
	Participants were recruited in two hospitals, a rehabilitation clinic and a physiotherapy practice in Bielefeld and Berlin, and at Bielefeld University. A purposive sampling approach was used to include three key groups of interest: patients (mental and physical conditions), social care users and carers (formal and informal).								
US	0	0	19	0	19	23-76 53.8 (13.8)	53	0.84 (0.20)α	77.3 (14.78)
	Respondents with acute and long-term physical and mental health conditions were recruited from clinics at the University of Illinois Hospital & Health Sciences System and the website ResearchMatch.org.								

N/C = not collected

† based on Devlin et al.[22]

§ based on Ludwig et al.[23]

α based on Pickard et al[24]

Table 2 Results of face validation studies

Item	UK	Argentina	Australia	China	Germany	US	Outcome (K/M/D)	Item taken forward
Domain: Activity								
I enjoyed what I did (F)	✓	✓	✓	Ø	✓	✓	K	
I was able to do the things I value (F)	✓	✓	✓	Ø	✓	✓	K	
I did things I found rewarding	×	✓	✓	Ø	✓	✓	D	
I was bored	×	✓	×	✓	✓	×	D	
I did what I wanted to do	×	✓	✓	Ø	Ø	✓	D	
I could do the things I wanted to do (F)	✓	✓	×	Ø	✓	✓	K	
I did what I needed to do	×	✓	✓	✓	✓	✓	D	
I was able to do what I needed (F)	✓	✓	×	Ø	Ø	×	K	

I had no difficulty with my day to day activities/daily activities	✓	Ø	×	Ø	Ø	Ø	M	How well were you able to do your day to day activities (e.g. working, shopping, travelling) (D)
Given the help I had/received my personal needs were met (e.g. being washed, going to the toilet, getting dressed, having food when I needed)	×	×	✓	Ø	Ø	✓	M	My personal needs were met (e.g. being washed, going to the toilet, getting dressed, having food when I needed) (F)
Given the help I had/received my self-care needs were met (e.g. being washed, going to the toilet, getting dressed, having food when I needed)	×	Ø	×	Ø	Ø	✓	D	
I was able to look after myself (F)	Ø	✓	✓	Ø	Ø	✓	K	
I needed help with looking after myself	×	✓	×	Ø	Ø	✓	D	
I was able to look after myself with no difficulty	Ø	Ø	✓	✓	Ø	Ø	M	I was able to look after myself (e.g. being washed, going to the toilet, getting dressed, having food when I needed) (F)

I had no difficulty with self-care activities	×	Ø	×	✓	Ø	Ø	D	
I was able to get around inside my home with no difficulty (D)	✓	✓	✓	✓	✓	Ø	K	
I was able to get around outside with no difficulty (D)	Ø	✓	✓	✓	Ø	Ø	K	
How well did you communicate with others	×	Ø	Ø	✓	Ø	✓	D	
I was able to communicate with others with no difficulty	×	Ø	×	✓	✓	Ø	D	
Because of hearing and/or speech, how difficult did you find it to have a conversation (D)	✓	Ø	×	Ø	Ø	✓	K	
How well can you hear (using hearing aids if needed)	✓	✓	✓	Ø	Ø	✓	M	How well can you hear (using hearing aids if you usually wear them) (D)

I had no difficulty hearing (using hearing aids if needed)	×	×	×	Ø	✓	Ø	D	
How well can you see (using your glasses or contact lenses if they are needed) (D)	✓	✓	✓	Ø	Ø	✓	K	
I had no difficulty seeing (using your glasses or contact lenses if they are needed)	×	×	×	Ø	✓	Ø	D	
								New item: I was able to do the things I wanted to do (S)
Domain: Autonomy								
I felt able to cope	✓	✓	✓	×	Ø	Ø	M	I felt able to cope with my day to day life (F)
I felt unable to cope	×	×	×	×	×	×	D	
I felt unable to cope with my day to day life (F)	✓	×	✓	Ø	Ø	Ø	K	

I felt overwhelmed by my problems	✓	✓	×	Ø	Ø	×	M	I felt overwhelmed by the problems or situation (F)
I felt in control of my daily life	×	Ø	×	Ø	Ø	✓	D	
I felt in control of my day to day life (F)	✓	✓	✓	Ø	✓	✓	K	
I have as much control over my daily life as I want	Ø	Ø	×	Ø	Ø	×	M	I had control over my day to day life (F)
								New item: I felt I had no control over my day to day life (F)
Domain: Cognition								
I found it hard to concentrate (F)	✓	✓	Ø	Ø	✓	✓	K	
I found it hard to focus my thoughts	×	×	×	×	Ø	Ø	D	
I found it hard to pay attention (F)	✓	✓	✓	✓	✓	✓	K	
I had trouble thinking clearly (F)	✓	✓	×	✓	Ø	×	K	
I had trouble remembering (F)	✓	✓	Ø	✓	✓	✓	K	
I had trouble with my memory	×	✓	✓	Ø	Ø	×	D	

I felt confused (F)	Ø	Ø	Ø	Ø	×	×	K	
Domain: Feelings and emotions								
I felt happy (F)	✓	✓	✓	Ø	✓	✓	K	
I felt unhappy (F)	✓	Ø	✓	Ø	✓	×	K	
I felt depressed	×	✓	✓	Ø	Ø	×	D	
I felt sad (F)	Ø	✓	✓	Ø	✓	✓	K	
I enjoyed life	×	Ø	✓	Ø	✓	✓	D	
I felt content with my life	×	✓	×	✓	×	✓	D	
I thought my life was not worth living (F)	✓	✓	×	Ø	×	×	K	
I felt that I had nothing to look forward to (F)	Ø	✓	✓	×	Ø	×	K	
I had nothing to look forward to	×	×	×	Ø	Ø	×	D	
I looked forward to each day	×	Ø	×	Ø	✓	✓	D	
I felt frightened (F)	Ø	×	✓	✓	✓	×	K	
I felt afraid (F)	Ø	✓	✓	✓	✓	✓	K	

I felt safe (F)	✓	Ø	✓	Ø	Ø	×	K	
I felt unsafe (F)	✓	×	✓	Ø	Ø	×	K	
I felt secure	×	×	✓	Ø	Ø	✓	D	
I felt anxious (F)	✓	Ø	✓	Ø	Ø	✓	K	
My worries overwhelmed me	×	✓	×	×	Ø	×	D	
I felt worried (F)	✓	✓	✓	Ø	✓	✓	K	
I felt calm (F)	✓	✓	✓	✓	✓	✓	K	
I felt relaxed	×	✓	×	✓	✓	×	D	
I felt irritable (F)	✓	Ø	✓	✓	✓	✓	K	
I felt irritated	×	Ø	✓	✓	✓	×	D	
I felt angry (F)	✓	✓	✓	✓	✓	✓	K	
I felt frustrated (F)	✓	✓	✓	✓	✓	✓	K	
I lost my temper easily (F)	✓	✓	×	✓	Ø	×	K	
								New item: I felt cheerful (F)
Domain: Physical Sensations								
I had no pain (mild pain etc.).	✓	✓	✓	Ø	✓	Ø	M	I had no physical pain (mild pain etc.) (S)

How often do you experience pain	✓	✓	✓	Ø	✓	Ø	M	How often do you experience physical pain (F)
I had no discomfort (mild discomfort etc.).	✓	✓	✓	Ø	Ø	Ø	M	I had no physical discomfort (mild discomfort etc.) (S)
How often do you experience discomfort	✓	✓	✓	Ø	Ø	Ø	M	How often do you experience physical discomfort (F)
I felt exhausted (F)	✓	✓	Ø	✓	✓	✓	K	
I got tired easily	✓	✓	×	✓	✓	✓	M	I felt very tired (F)
I was too tired to do anything	×	✓	×	✓	✓	×	D	
I had problems with my sleep (F)	✓	✓	✓	✓	✓	✓	K	
Domain: Relationships								
I felt supported by other people	×	Ø	Ø	Ø	✓	✓	D	
I felt unsupported (F)	✓	✓	Ø	✓	Ø	Ø	M	I felt unsupported by people (F)
Other people gave me support	×	Ø	×	Ø	✓	×	D	
I had support when I needed it (F)	✓	✓	✓	✓	✓	✓	K	

I had disagreements and conflict with people	×	×	×	Ø	✓	×	D	
I got on with people around me	✓	Ø	Ø	×	Ø	Ø	M	I got along well with people around me (F)
I got along well with people I came into contact with	×	Ø	×	✓	Ø	Ø	D	
I felt lonely (F)	✓	✓	✓	✓	✓	✓	K	
I felt there was nobody I was close to (F)	Ø	✓	×	✓	×	✓	K	
I felt I had no one to talk to (F)	✓	✓	×	Ø	Ø	✓	K	
I felt isolated (F)	Ø	Ø	✓	✓	Ø	×	K	
I felt people avoided me (F)	Ø	✓	✓	✓	✓	×	K	
I felt judged by others	×	×	×	✓	Ø	×	D	
I felt accepted by others (F)	✓	✓	✓	✓	✓	✓	K	
I felt excluded (F)	✓	✓	✓	✓	✓	Ø	K	
I felt left out (F)	✓	✓	✓	✓	✓	✓	K	
Domain: Self-identity								

I felt confident in myself (F)	✓	✓	✓	Ø	Ø	✓	K	
I felt confident	×	Ø	×	Ø	✓	×	D	
I felt unsure about myself (F)	✓	✓	✓	✓	×	×	K	
I felt I was treated with respect	×	✓	✓	✓	✓	×	D	
I felt respected	×	✓	×	✓	Ø	×	D	
I felt like I lived with dignity	×	×	×	Ø	×	×	D	
I felt good about myself (F)	✓	✓	✓	Ø	Ø	✓	K	
I felt like a failure (F)	Ø	✓	✓	✓	×	×	K	
I felt valued	×	✓	×	✓	Ø	×	D	
I felt useful	×	✓	×	Ø	Ø	✓	D	

✓ no problems identified; × problems identified; Ø mixed evidence

K=Keep; M=Modify; D=Drop; F=Frequency response option; S=Severity response option