



HHS Public Access

Author manuscript

J Appl Meas. Author manuscript; available in PMC 2019 April 02.

Published in final edited form as:

J Appl Meas. 2010 ; 11(3): 315–330.

Opportunities for CAT Applications in Medical Rehabilitation: Development of Targeted Item Banks

Pamela Kisala, M.A. and

University of Michigan, Ann Arbor, MI

David S. Tulsy, Ph.D.

University of Michigan, Ann Arbor, MI

Abstract

Researchers in the field of rehabilitation medicine have increasingly turned to qualitative data collection methods to better understand the experience of living with a disability. In rehabilitation psychology, these techniques are embodied by participatory action research (PAR; Hall 1981; White, Suchowierska, & Campbell 2004), whereby researchers garner qualitative feedback from key stakeholders such as patients and physicians. Glaser and Strauss(1967) and, later, Strauss and Corbin (1998) have outlined a systematic method of gathering and analyzing qualitative data to ensure that results are conceptually grounded to the population of interest. This type of analysis yields a set of interrelated concepts (“codes”) to describe the phenomenon of interest. Using this data, however, becomes somewhat of a methodological problem. While this data is often used to describe phenomena of interest, it is challenging to transform the knowledge gained into practical data to inform research and clinical practice. In the case of developing patient-reported outcomes (PRO) measures for use in a rehabilitation population, it is difficult to make sense of the qualitative analysis results. Qualitative feedback tends to be open-ended and free-flowing, not conforming to any traditional data analysis methodology. Researchers involved in measure development need a practical way to quantify the qualitative feedback. This manuscript will focus on a detailed methodology of empiricizing qualitative data for practical application, in the context of developing targeted, rehabilitation-specific PRO measures within a larger, more generic PRO measurement system.

The 21st century has seen several large scale federally sponsored initiatives to develop new state-of-the-art comprehensive, multidimensional outcomes measures for use in clinical trials research. Advanced by the Director of the National Institutes of Health (NIH), the 2002 NIH “roadmap” initiative was designed to improve the research infrastructure and thereby increase the capacity of the biomedical research enterprise (Division of Program Coordination, 2009; Lou Quatrano et al Special Issue 2010?). A key component of the

Correspondence concerning this article should be addressed to Pamela A. Kisala or David S. Tulsy, Kessler Foundation Research Center, 1199 Pleasant Valley Way, West Orange, New Jersey 07052. pkisala@kesslerfoundation.org, pamela.kisala@yahoo.com, dtulsy@kesslerfoundation.org, or dtulsy@verizon.net.

Author Note

Pamela A. Kisala, Spinal Cord Injury Research Laboratory, and Outcomes and Assessment Research Laboratory, Kessler Foundation Research Center; David S. Tulsy, Spinal Cord Injury Research Laboratory, and Outcomes and Assessment Research Laboratory, Kessler Foundation Research Center, and University of Medicine and Dentistry of New Jersey.

roadmap prioritizes the “re-engineering” of outcomes measure development and a flagship project which involves the development and validation of the dynamic assessment of Patient Reported Outcomes (PROs) for clinical trials research. Since 2002, the NIH has sponsored large initiatives to develop measurement tools for use across all of their patient populations. This includes the Patient Reported Outcomes Measurement Information System (PROMIS) (Cella, Yount, et al. 2007; Division of Program Coordination, 2009), the Neuro-QOL measure for individuals with neurological disorders (“Quality of Life,” 2009), and the Toolbox for Neurological and Behavioral Functioning (“NIH Toolbox,” 2009; Cella, Gershon, Lai, & Choi, 2007). The resulting tools have been developed using state of the art measurement theory and methodology including item banking(Choppin, 1968, 1976, 1978, 1979, 1981), Item Response Theory (IRT) (Hays, Morales, & Reise, 2000), and Computerized Adaptive Testing (CAT; Cella, Gershon, et al., 2007). These new measures, groundbreaking in their scope and levels of federal funding, utilize advanced psychometric techniques and employ state-of-the-art administration methods. Therefore, these instruments promise to reshape biomedical outcomes measurement in the 21st century.

Often, these new instruments are generic in nature and designed to be used across diverse patient populations. These instruments therefore form a universal measurement system that can be used across studies. While such measures contain items that are common across populations, the scales do not address any issues which may be unique to one’s specific condition or disease. For example, a sudden and traumatic injury, such as spinal cord injury (SCI), changes an individual’s world drastically and forever in an instant. Additional condition-specific items need to be written and included in the larger measurement systems to ensure targeted measurement of relevant issues and thereby adequately evaluate important outcomes in rehabilitation populations. As rehabilitation researchers face an unprecedented growth in the number of clinical trials designed to restore functioning and improve health-related quality of life (HRQOL) in individuals who have sustained a traumatic injury, the need for targeted measures becomes more important than ever. The paper will provide an overview of how qualitative analyses can be employed to enhance the content validity and ultimately the clinical sensitivity of a scale. Specifically, this paper will illustrate the expansion of such measurement efforts into a rehabilitation population, describing how patient feedback can guide item bank development so that more sensitive instruments can be developed.

Item Banking and CAT

At the outset of the item pool development process, it is imperative to clearly and precisely define and refine the topic of interest (Spector, 1992), clarifying the scope of the ultimate scale. The process of scale development begins in earnest with generation of a large pool of items to represent the single construct to be measured. In recent years, IRT has paved the way for practical implementation of this new generation of outcomes measurement tools in both research and clinical settings (Flaughner, 1990). Many new generation scales such as PROMIS and Neuro-QOL are intended for use with an IRT / CAT platform. Hays et al. (2000) outline the many advantages of adopting IRT methodology in the field of health outcomes measurement, such as independent assessment of latent trait and item difficulty, respectively, assessment of differential item functioning (DIF), and ability to detect change

over time. To successfully carry out this type of psychometric analysis, steps must be taken to ensure that several key assumptions underlying IRT are met. For example, each domain must measure a single construct. Any items that are not deemed to measure this single underlying concept are removed from the item pool. In the case of a multidimensional construct, it is necessary to break this down into a series of a priori domains and perhaps subdomains. Each domain or subdomain of content then forms the basis of an individual, unidimensional item pool within the larger multidimensional construct or system of measures. It is also important to have enough items in each preliminary item pool to ensure that a sufficient number of calibrated items will be evenly distributed across the continuum of severity within the given subdomain. Each final item pool is then tested in a large scale calibration sample (e.g. n = 200 if following a 1 parameter logistic (PL), or Rasch model, n = 500 for a 2-PL model; Hambleton, 1989; Linacre, Heinemann, Wright, Granger, & Hamilton, 1994). Factor analysis of collected data is used to confirm unidimensionality of an item pool, thereby confirming or refuting the a priori subdomain decisions made by the project team. Subsequently, IRT analyses (Hill et al., 2009; Reeve et al., 2007; Hays, Liu, Spritzer, & Cella, 2007) are used to select items for inclusion/exclusion in the final item bank based on item calibrations and fit statistics, and to calibrate the included items on a single underlying metric. It is important to note that a group of items is a “pool” prior to this psychometric analysis, and only the final calibrated items may carry the title of item “bank”. (Cella et al 2007; Choppin 1976; Choppin 1979; Ware, 2004) Once IRT is used to relate items to one another in this way, the researcher may now predict the probability of an individual responding a certain way to a specific item given that individual’s response to a previous item, to develop calibrated short forms which use the most informative items to achieve satisfactory standard error while minimizing participant and administration burden (Cella, Gershon, et al., 2007), and, ideally, may use the psychometric data to inform CAT measure development. The use of CAT eases administration burden while enabling evaluators to administer succinct, discrete item sets that exhibit solid psychometric accuracy, as well as a greater degree of discrimination, than fixed-length assessments (Cella, Gershon, et al., 2007). CAT allows researchers to test the full range of each construct with just a few algorithm-selected items, increasing precision across the entire range of the construct while limiting respondent burden. The resulting increase in sensitivity and specificity of the final scale improves both research and clinical utility of the scale. This paper will serve as a guide to identifying the relevant content for inclusion in the item pool, and to demonstrate a practical method to integrate qualitative feedback into development of an item pool. Other papers in this series will cover the use of IRT to develop the item pool into a calibrated item bank.

Identification of Relevant Content

Items in the majority of existing measures have been developed using general health populations and some of these items may be irrelevant or even inappropriate for individuals with disabilities (Andresen & Meyers 2000). As such, it is clear that improvements could be made when developing rehabilitation outcomes instruments. A first step in the development of IRT-enabled PRO measure development in a rehabilitation population is to determine the issues that are important to the target population and write a large, comprehensive pool

of relevant items. Of paramount importance is to identify relevant subject matter, or the specific domains and subdomains of content, for inclusion. Given the absence of a “blueprint” (Bode, Lai, Cella, & Heinemann, 2003) to determine precise constructs for inclusion in HRQOL measurement, researchers, consumers, and clinicians must work collaboratively to identify conceptually grounded subject matter for inclusion. Each of the evolving PRO measurement systems has utilized expert input from a variety of important disciplines, including measurement experts, leading research scientists, and content experts in areas as diverse, yet complementary, as cognitive science, technology, qualitative research, and psychometrics. Additionally, the elicitation, interpretation, and inclusion of key stakeholder perspectives is the cornerstone of Participatory Action Research (PAR; Hall, 1981; White et al., 2004), a methodology whereby key stakeholders (e.g. physicians/clinicians, individuals with the disease or disorder in question and their family, friends, caregivers) work directly with researchers in an iterative manner to collaboratively explore components and determine key indicators of each construct in question. Following a systematic review of the PAR literature, White et al. (2004) highlight increased content relevance (“social validity,” p. 1) and improved consumer accessibility of final instruments as advantages of this grounded, yet rigorous methodology. Additional rigor is added through iterative expert review and revision of the domain framework and, ultimately, selection of specific subdomains for inclusion. This step by step process, outlined in Figure 1, is described in greater detail below.

Once the investigative team has identified the most important subject matter for inclusion in the new item pool(s), the team’s focus may then shift to writing items (see Diagram 2). The first step in item pool development is the comprehensive review of existing measures. Bode et al. (2003) provide a systematic framework for equating existing items along a common metric. The selected items from existing measures form the initial basis of the new HRQOL item pool. Subsequently, specific consumer statements, issues, and concepts within each selected subdomain are transformed into PRO items. Consumer-based items are supplemented by items developed by subject matter experts (i.e., individual(s) with a history of research and publication within the given domain/subdomain area). As with the domain hierarchy itself, the initial item pool is subject to several rounds of iterative expert review and revision, as well as cognitive debriefing with consumers, to ensure the continued grounding of item pool content throughout the item development process. This iterative interaction of researcher and community stakeholder is also referred to as “community based” partnership research. (Israel Schulz, Parker, & Becker, 1998) This community / consumer participation ensures that the final measure and ultimately the data are relevant and significant at the consumer level. The resultant item pool, when sampled appropriately, yields highly reliable measurement at all points on the continuum of health or functioning in a given domain (Revicki & Cella, 1997).

Participatory Action Research

Hall (1981) coined the term “participatory action research) and White, Suchowierska, and Campbell (2004) have reviewed initial and subsequent definitions of the term. In essence, PAR consists of stakeholders such as patients and providers forming a partnership with members of the research team, with patients or consumers and researchers alike providing

input to one another throughout the course of the project. In this way, the project team will ensure that the final scale is conceptually grounded in that content which is most important to patients, providers, and other key stakeholders. In the initial stages of PAR, researchers must first determine who to include in their sample, and then must decide how they will open a dialogue to ascertain the most relevant information. Once a determination is made regarding which stakeholders are most important to include, it is necessary to select a methodology for collection of qualitative feedback from these individuals. Focus groups, individual interviews, and paper and-pencil surveys are common means of ascertaining participant feedback. The authors will refer to focus groups throughout the text, however individual interviews and paper-and-pencil surveys are other common means of garnering stakeholder feedback.

Data Collection

In anticipation of participant recruitment and to avoid obtaining a sample of convenience, investigators should ideally examine overall demographic characteristics of the population in question to obtain a representative sample with regard to key factors such as age, race, and condition-specific variables. A sufficient number of groups should be scheduled to ensure a multifaceted stakeholder voice distinct from the researchers' preconceived notions. The research team should employ a systematic approach to the development of detailed guides for the focus group moderator(s). Moderators must be neutral and not dictate the direction of the discussion. Of primary importance in guide development is determination of goals of the groups. Do the researchers want to generate domains, confirm existing domains, generate new items, review existing items, or a combination of the above? Focus group guides used in other research may be adapted and refined for the current purpose. Additionally, it may be useful to invite other researchers experienced in focus group facilitation to serve as expert consultants to the focus group guide development process. To facilitate qualitative analysis of focus group feedback, it is helpful to obtain verbatim transcripts of focus group proceedings. This may be done by audio- or video-recording and transcribing (either in house or through an outside vendor), or by hiring a court reporter to transcribe the proceedings in real time.

Several key steps must take place to transform open-ended qualitative data into useable, measureable information. One potential limitation of utilizing qualitative data is potential for investigator bias to selectively emphasize some information, while ignoring information that does not conform to ones preexisting beliefs. The methodology described below can be used to reduce this inherent bias in analysis of focus group data. Glaser and Strauss (1967) warn against the tendency of "examplifying" or deriving "dreamed-up, speculative, or logically deduced theory," (p. 5). The process of coding and reconciliation ensures focus and attention go to new or unpopular ideas, those that may be more distant from the researcher's frame of reference. Such ideas, once identified, are then evaluated to determine their relative merit for inclusion in a questionnaire. This iterative analytical process ensures optimal conceptual grounding and generation of a substantial number of targeted, condition-specific issues. This method ensures that stakeholder feedback will be fully evaluated, even in cases where such feedback may differ from the researchers' preconceived hypotheses.

Qualitative Analysis

Grounded theory based qualitative analysis (Glaser & Strauss, 1967);Strauss & Corbin, 1998) may be used to transform open ended information into more measurable empirical data, and thereby serves as a mechanism to incorporate the results of PAR directly into outcomes measure development. Much more than a cursory content analysis, the research team thus implements a systematic way of extracting and quantifying relevant themes discussed in stakeholder groups. The qualitative (e.g. focus group) data serves two concurrent purposes: 1) To construct a conceptually relevant hierarchical framework to serve as a basis for the frequency analysis of themes and and 2) To develop a comprehensive item pool of issues targeted to the population in question. Relative frequencies subsequently influence weighting (frequency) of items in each topic area.

For the purpose of domain development, the initial goal of qualitative analysis is to discover specifically what overarching concepts and content domains are important to members of the population in question. The qualitative analysis should be structured around a systematic review of data, to minimize examiner bias and subjectivity in review of the data and development of item bank content. Detailed information about the qualitative feedback is aggregated in the form of frequency data which helps to guide item bank development, ensuring appropriate and adequate attention is paid to areas cited as important to patients and other key stakeholders. The second and equally important goal of the qualitative analysis is to generate data: that is, to brainstorm, to develop as many new issues and ideas related to the construct in question as possible. The method places the patient in the role of expert in the item development process. Relevant issues become the basis for individual items. These items then combine to form a first draft of the new, targeted, condition specific item pool.

Domain Development Process

The purpose of domain development is to quantify the qualitative data. The quantification is done by determining the frequency with which participants have generated issues and provided feedback on various qualitative topics to the researcher. Domain development contains Open, Axial, and Selective coding processes (see Figure 1). Open coding is the initial review of the data, where the researcher identifies themes as they emerge. Axial coding is finding the relationships between these themes; the structure could be relational, hierarchical, or clustered, and is typically performed by the PI or a research team member with considerable knowledge of the topic area. The deliverable product at the close of open and axial coding is the map (or “codebook”) for the raters to the transcript text. This final stage, Selective coding, consists of having individual raters go through and code each line of transcript text according to the codebook developed by the research team. The final result will be one code for each line of text. This process will be described in more detail below.

Open and Axial Coding

For each domain, two or three independent investigators identify major content areas and develop an initial list of subdomains (“Codes”) through thorough transcript review (a process known as “open coding” (Strauss &Corbin, 1998)). These broad, conceptual level

codes provided an initial structure for identifying themes in the focus group transcripts. Transcripts of all focus groups are reviewed, and codes are related to one another to form a hierarchical taxonomy – or “codebook” (MacQueen, McLellan, Kay, & Milstein, 1998) for each domain. Known as Axial Coding (Strauss & Corbin) or “focused coding,” (Charmaz, 1995) this iterative process consists of multiple reviews, investigator teleconferences, and revisions. Verbatim focus group quotes are recorded to support development of new codes. During coding, an iterative, constant comparative process is used to identify new themes as they emerge. A literature review and a review of current instruments should be conducted to further inform codebook development. Each domain is then refined and expanded to reflect the emerging themes. Next, focus group content themes and the domain coding structure are independently reviewed and confirmed by other researchers with specific expertise in each domain area. Ultimately, a detailed, all-inclusive and mutually exclusive (Gorden, 1992) codebook emerges for each domain, with examples, definitions, inclusion criteria, and exclusion criteria added to clarify and semantically anchor the meaning of each code.

Selective Coding and Descriptive Analysis

The next step in the qualitative analysis process involves “Selectively” coding all transcripts to identify all instances of all codes included in the codebook. Carey, Morgan, and Oxtoby (1996) emphasize the importance of pre-testing the codebook to ensure that two independent raters are able to consistently interpret the codebook given the same coding instructions. For each domain, it is imperative that the raters review each code together to ensure consistent understanding at all levels of specificity within the codebook. Raters must be anchored in interpretation of codes to ensure the highest possible inter rater reliability in the application of codes to transcript text. To assess rater synchronization prior to application of codes, raters independently complete a preliminary inter-rater reliability exercise to standardize their use of the codebook. A method of doing this would be to create an exercise consisting of 25 verbatim statements from focus group transcripts to which the raters are to assign the appropriate code. The three raters must achieve at least 80% overall agreement to progress to the next step. If raters do not achieve a minimum of 80% agreement on the first exercise, codebook definitions must be expanded to enhance clarity and a second exercise is completed, with this process continuing iteratively until agreement reaches the satisfactory level. In every case of overall agreement <100%, raters must review all discrepancies to determine if discordance is due to 1) lack of mutual exclusivity or 2), lack or rater synchronization. In the case of the former, codes may be merged, or code definitions further refined to a point of mutual exclusion. If raters find the latter to be the source of inconsistency, they must, at the very least, agree (with a third party if necessary) on the final code selection, and/or may add specificity to code inclusion/exclusion criteria to clarify circumstances for application of the code in question.

In the first attempt to apply this methodology, the researchers’ preliminary coding with a qualitative analysis software program (e.g. NVivo, 2008, ATLAS) yielded a significant barrier to determining consistency between coders as the indeterminate start and end points of each section of coded text became a formidable obstacle to coding comparison. Consequently, all data was carefully prepared to facilitate frequency analysis. Transcript text was divided into discrete, mutually exclusive “chunks” (Hoffman, Neville, & Fowlkes,

2006), the smallest units of dialogue related to a single topic or code; chunk size ranges from a single word to multiple sentences. To offset the inherent subjectivity of this exercise, this co-investigator did not make any determinations regarding relevance of each chunk; rather, raters coded the chunks exhaustively, placing items into a “Not Applicable to [Domain]” node where appropriate. To maximize rater synchronization and anchor code interpretation, the two raters completed “consensus coding” of the first transcript within each domain. The team hypothesized that with sufficient demonstrated inter-rater reliability, a single rater could code the remaining transcripts while maintaining relative accuracy in the calculation and reporting of descriptive statistics. After coding a second transcript independently, both raters’ coding was merged to facilitate computation of inter-rater reliability statistics. Percentage agreement for an independently-coded transcript was calculated as a simple percentage of chunks agreed upon divided by total chunks. Cohen’s kappa was calculated using the “Coding Comparison Query” feature in NVivo (2008). However, due to the inherent subjectivity of coding qualitative data, inter-rater agreement is bound to be lower than desirable to have only one person code each of the remaining transcripts. The best Cohen’s kappa values one could hope for would be in the range of .6-.7; while such values are generally considered to be “substantial agreement,” (Landis & Koch, 1977) the authors felt this allowed for too great a degree of subjectivity and inter-rater disagreement. Consequently, to reduce the coder bias in descriptive analyses, the research team should have two independent raters code each of the remaining transcripts, with a third individual responsible for reconciling discrepancies between the two raters. This is a more consensus based approach designed to minimize the inherent subjectivity of an individual rater applying codes independently. In other words, two raters independently rate each transcript. When complete, the raters log and reconcile with a third party all instances of disagreement. When completed for each transcript, this consensus process attempts to compensate for individual subjectivity and ultimately results in a single code being applied to each segment of transcript text. This will allow descriptive statistics to be calculated for each domain which will provide frequency data to inform domain structure and item development decisions.

Information Synthesis

Following the detailed qualitative analysis process described above, the research team is left with the task of prioritizing and synthesizing the wealth of resultant codes, frequency data, and expert input. The relative frequency of mention of each subdomain is an important piece of information considered in selecting subdomains for inclusion in the final measure. However, other factors must be considered as well including the theory or conceptual framework of the measure, or issues that have a high valance but might only impact a small sector of the population. For example, in spinal cord injury (SCI), respiratory complications have serious, if not fatal, effects on QOL in a subgroup of the population – namely especially among those individuals with high level cervical injuries. While respiratory complications impact a small percentage of the SCI population, the morbidity and mortality of these problems could warrant inclusion in the measure. Additionally, participants are likely to discuss avoid topics of conversation that are not generally discussed in public, which may inhibit discussion of sensitive topics such as sexuality especially if the feedback

is obtained in a group setting. Focus groups are generally small and researchers need to ensure that there isn't a sampling bias preventing the researcher from obtaining feedback from harder to reach segments of the target population (e.g., those most likely to attend focus groups are those who possess sufficient physical ability, time, and transportation to attend focus groups. Such biases could result in a restricted range of responses, and expert input and current literature must be considered in conjunction with focus group data to ensure that the entire patient population in question is represented. Also, because of the subjective nature of the data, involvement of multiple people within the investigative team, in reviewing and decision making, is important. Once the final domains and subdomains are selected for development, then the transcripts are re-reviewed to pull additional content and write items. Many new items are drawn directly from language used in groups.

An Illustration

The researchers received NIH funding (R01 HD054659) develop a new PRO measure system targeted for individuals with SCI that is integrated with the PROMIS and Neuro-QOL measurement systems. This project, called the SCI-QOL (Tulsy, Kisala, Victorson et al, 2010), will serve as a case example to demonstrate a practical method for conducting qualitative analysis and integrating qualitative feedback into the development of an item pool. In this case example, the research team set out to develop a measure of a health-related quality of life (HRQOL) for use with an SCI population. The initial conceptualization of the construct in question was drawn from Cella's (1995) definition of HRQOL as "The effects a medical condition or its treatment has on one's usual or expected physical, emotional, and social well being" (p. 73). An assessment of HRQOL therefore enables evaluation of meaningful change from the patient's perspective. Existing , generic measures of HRQOL [e.g. the Medical Outcomes Study short form (SF-36; Ware & Sherbourne, 1992), Sickness Impact Profile short form (SIP68; de Bruin, 1997)] often do not contain relevant domain content when used with a targeted population, and in fact are unable to measure HRQOL in disability populations without some potential biases such as inappropriate wording (Andresen & Meyers, 2000). To build a new measure that is more relevant and sensitive for use in an SCI population, the researchers began with the Neuro-QOL conceptual framework of HRQOL. Following a literature search of existing measures, extensive stakeholder feedback was incorporated to see how the condition in question – SCI – was underrepresented in not only more traditional measures of HRQOL, but also in the newer, more advanced generic measures. This qualitative data served to confirm or refute the importance of each Neuro-QOL subdomain, to supplement the Neuro-QOL conceptual structure with SCI-specific areas of interest, and to flesh out the detailed subdomains of HRQOL that are particularly relevant for this population. The research team followed PROMIS and Neuro-QOL definitions and methodology, using PROMIS and Neuro-QOL documents as templates for all SCI-QOL study materials.

During the initial data collection phase of the study, researchers conducted focus groups at four Spinal Cord Injury Model Systems sites across the country to define specific areas in need of measurement. Individuals with spinal cord injury were the primary stakeholders of interest, and clinical care providers (eg physiatrists, nurses, psychologists, physical therapists) were secondary stakeholders who would form a community based partnership

with the researchers to determine the most important and relevant issues for inclusion. Targeted recruitment yielded a representative SCI sample with regard to key demographic characteristics including gender, race, and level of injury (National Spinal Cord Injury Statistical Center, 2007, 2009).

Sixty-five individuals with SCI and 42 SCI clinicians participated in a total of 16 focus groups which were designed to serve two concurrent purposes. First, it was imperative to ensure that no SCI-specific domains or subdomains of content were missing from the conceptual structure set forth by the Neuro-QOL project team (see Table 1). The second goal of the groups was to produce a list of potential items for inclusion in the HRQOL item pool. The research team emulated and refined the Neuro-QOL template for conducting semi-structured yet open-ended focus groups, making minor changes in an effort to elicit item-level feedback from participants. Participants were asked to spontaneously generate relevant HRQOL domains and issues. Once this discussion concluded, participants were asked to delve more deeply into a specific domain of HRQOL (e.g. Physical-Medical Health, Emotional Health, Social Participation) and again spontaneously generate relevant issues and, most importantly, potential items. Finally, participants were asked to review domains and issues from existing measures, comment on their relevance to SCI, and add any domains or issues they felt were missing. In this manner, the research team was able to elicit stakeholder feedback for direct incorporation into targeted item pools. All focus groups were audio-recorded and transcribed verbatim to facilitate analysis.

Focus group transcripts and NVivo 8 (QSR International 2008) software were used to conduct the systematic qualitative analysis procedure described above. Three separate analyses were conducted, one each for the Physical-Medical, Emotional, and Social Participation domains, respectively. The emotional domain analysis will serve as an example of the codebook development and coding processes. For this domain, three independent raters reviewed the entire set of Emotional transcripts (n=8; 4 patient group transcripts and 4 provider group transcripts). During the initial read-through, the raters each conducted “open coding” of the transcripts, compiling an ongoing list of issues/themes (“codes”) generated during the focus group discussions. Each rater reviewed each transcript twice to ensure all relevant themes were identified. Following this independent theme generation, the three raters met to compare codes. A master list was compiled, with raters agreeing on common nomenclature in instances where two raters had used different terms for the same semantic issue. All themes mentioned by at least one rater were included on this master list. The next step was “axial coding”, a process by which each rater independently related the codes to one another. Each rater thus created a draft “codebook”, a hierarchical structure of relationships between codes. The raters met again, comparing the three codebooks and discussing any structural discrepancies. After one rater combined all of the codebooks and feedback, the three raters continued to meet and make iterative revisions to the codebook until all raters were in agreement that the codebook accurately reflected the content and structure of the focus group data. At this point, the emotional domain consisted of 95 codes across 6 subdomains, namely “Emotions”, “Self Evaluation”, “Emotional Skills”, “Intimacy/Sexuality”, “Independence/Autonomy”, and “Emotional Roadblocks”. The raters collaborative created definitions and inclusion/exclusion criteria for each of the codes. The emotional codebook was then reviewed by an expert team of rehabilitation psychologists.

Expert suggestions, based on clinical expertise and familiarity with the current literature, called for expansion of subdomains. Specifically, experts documented a need for increased focus on the grief and loss often associated with SCI (Niemeier, Kennedy, McKinley, & Cifu, 2004), as well as on the resilience characteristic of individuals with positive psychological outcomes following SCI (Kennedy et al., 2009). The codebook was revised accordingly and approved by all project co-investigators. The next phase, selective coding, began with three raters (i.e., the two domain raters and a third individual who was instrumental to codebook development) independently completing the above-mentioned interrater reliability exercise to pretest the codebook and to anchor code interpretation and application. For this domain initial agreement was 88%. The raters reviewed and resolved all discrepancies, and finalized the codebook in NVivo. To further reduce subjectivity in utilizing the codebook, two raters coded the first transcript together, agreeing on one code for each segment of text. Any instances of disagreement were resolved by a third party. For each of the remaining transcripts, the two raters coded the transcript text exhaustively. NVivo was used to merge files and compare coding, and all instances of disagreement were itemized and resolved. In this manner, the raters achieved 100% agreement, resulting in a final file for each transcript with only one code for each chunk of text. These final transcripts were used to calculate descriptive statistics.

Resulting frequency data helped researchers to flag key relevant content and, ultimately, to guide domain and subdomain selection for the final measure (Tulsy et al., 2010). A collaborative team of expert co-investigators considered relative frequency of mention of each subdomain in conjunction with a current review of the literature and iterative expert input to determine which domains/subdomains to include and which to omit. Some of the more generic Neuro-QOL and PROMIS domains such as Depression and Social Role Satisfaction were confirmed as important to individuals with SCI, while the Neuro-QOL domain of Cognitive Health, for example, was not deemed to be an SCI-related PRO. Perhaps most notably, several new, SCI-specific subdomains emerged. Please see Table 2 for a complete list of SCI-QOL domains which includes linkages to Neuro-QOL and PROMIS. A detailed description of the population-specific methods (eg moderator prompts, sample size and characteristics) utilized in the item development stage of the Spinal Cord Injury Quality of Life (SCI-QOL) measure is reported elsewhere (Tulsy et al., 2010).

Item Bank Development

Investigators must review existing measures, and select any extant items that will serve as a basis for the new item pools. Such inclusion of items from other measures may also facilitate the psychometric linking (e.g. predicting, aligning, or equating) of scores on one measure to scores on another (Brennan, 2006). Concurrently with codebook development, investigators maintain an ongoing log of relevant issues and ideas raised by focus group participants, greatly expanding the draft item pool. Building on item pool development techniques set forth by the PROMIS and Neuro-QOL teams, this overall pool of thousands of seemingly disparate ideas and issues evolves into a small set of discrete, unidimensional item pools (see Figure 2). First, each issue is reworded as an item (see Table 3) following the general PROMIS / Neuro-QOL item format. This initial list of items is merged with the extant items for inclusion, and investigators independently grouped similar responses together into “bins”

according to domain and subdomain) to facilitate further review. Investigators independently confirmed that all items were properly “binned” or systematically grouped by underlying latent trait (DeWalt, Rothrock, Yount, & Stone, 2007). In essence, each item was assigned to the appropriate code in the codebook generated during qualitative analysis. At this time, the grouped responses were reviewed to confirm the current codebook structure (i.e. node hierarchy) and to assess the need for generation of new domains/subdomains or even additional items in existing subdomains.

General items and those redundant with extant items are deleted in an effort to defer to the verbatim extant items wherever possible. Remaining items with similar content are organized along a unidimensional item hierarchy, with the “negative” items grouped at one extreme and the “positive” items grouped at the other. It is likely that the most “negative” items will be endorsed only by those individuals experiencing the greatest degree of distress or disruption and not at all by other individuals. For example, the item, “My fatigue prevents me from getting out of bed,” is one of the most extreme items within a potential fatigue domain that will probably be endorsed by only the *most fatigued* individuals. Alternatively, an item at the opposite extreme of the item hierarchy, “My high energy level allows me to work 60-hour work weeks,” will probably be endorsed only by the *most energetic* people. The team conducts a preliminary “winnowing” (DeWalt et al., 2007) of the item pools once the item hierarchies have been established, evaluating any items that appear redundant. Each item is compared to other items within its domain to determine whether it addresses a unique dimension, or conversely, if it is so similar in content to another item that it is simply repetitive. Only the most clear, comprehensive, and well-written item of a conceptually redundant group of items should be retained. Items are eliminated if they appear inconsistent with the domain definition or if they are poorly worded (e.g., vague, confusing, or narrow), and all items that are redundant or irrelevant are moved to a “deleted items” bin. Retained items that are poorly worded (e.g. vague, confusing, or narrow) or reflect multiple concepts are reworded as appropriate to improve clarity and conciseness, and continue to be included in the current item pool for each domain. The authors edited all items to ensure that item format was consistent with PROMIS and Neuro-QOL stylistic conventions. Following this process of item elimination and revision, all of the remaining items are reviewed to ensure appropriateness, ease of reading, and consistency with the domain definition.

Gaps exist in an item hierarchy when the items do not reflect all levels of the construct that they are intended to measure. Members of the investigative team use a systematic methodology to review the existing items and assess and resolve instances of construct deficiency (i.e., when there are not enough items to adequately measure the construct across the entire continuum). New items are then developed in each case of inadequate construct representativeness. The formation of additional items should adhere to at least one of the following guidelines: a) utilization of an “item writing” committee, comprised of clinical and statistical experts, b) use of community feedback (through focus groups and individual interviews) in the creation of new items, or c) use of condition-specific items, developed during prior or pilot studies, that have yet to be field tested.

Once the draft item pool is complete, qualitative item review (QIR; DeWalt et al., 2007) is used to “winnow” down the item pool in each domain to the desired number of items for the

purpose at hand. In an effort to ensure the developing item pools are consistent with the current state of the art, all new items are reviewed by senior scientists who are each leading domain experts specifically with regards to item pool development and/or the patient population in question. Expert reviewers are asked to write, revise, or delete items as appropriate to ensure that the final result is a pool of items which was comprehensive with regard to subject matter, clear, precise, acceptable to respondents, and appropriate for the planned data collection and analysis (Cella et al 2006). This expert review and revision serves to incorporate new thoughts and ideas on the forefront of science in each domain; this phase of review also allows for the assimilation of issues and ideas from concurrent work by other research teams. The second phase of QIR consists of cognitive debriefing interviews (Willis, 1999) with members of the target population to gather information on item comprehension, retrieval from memory of relevant information, and decision and response processes (Tourangeau, 1984). A minimum of 5–10 individuals from the target population should review each item and provided feedback on item context, time frame, and wording; discuss how they arrived at each answer; and rephrase items in their own words. Participants are asked to flag any inappropriate items and are given an opportunity to suggest additional items they feel are “missing”. This cognitive debriefing methodology ensures the conceptual grounding of the final item pool to issues of subjective importance to primary stakeholders. Cognitive interview feedback is reviewed by a minimum of two independent investigators, and is incorporated into the item pools when deemed appropriate. Subsequently, to ensure amenability to linguistic translation, a team of language translation experts evaluate each item for feasibility of translation. Items, words, or phrases that would be potentially problematic are identified, and are modified whenever possible. To ensure maximum utility of the final scale, the Lexile Framework for Reading (MetaMetrics, 1995) is used to confirm that all items fall at or below a sixth-grade reading level. After any necessary modifications, the final item pool is then re-distributed to all collaborating investigators for any final modifications prior to field testing. The specifics (e.g. item counts, cognitive interview feedback) of the SCI-QOL item development process are enumerated elsewhere (Tulsy et al., 2010).

At this point, each item pool is ready for large-scale (i.e. $n \geq 500$) calibration testing. Psychometric performance will inform final item candidates for inclusion in the calibrated item bank for each domain. Specifically, IRT analyses will detect poorly functioning items, determine the difficulty level and discrimination of each item, determine the probability that a particular individual will respond to an item in a certain way given their overall level of QOL, and finally, flag items that appear to perform differently in different population subgroups – for example men and women – an undesirable phenomenon known as differential item functioning (DIF).

Discussion

The authors have presented a systematic method for eliciting and incorporating stakeholder feedback in the development of targeted patient reported outcomes measure that is appropriate for implementation across a wide range of patient populations. This work may serve as a blueprint for the development of targeted item banks for other CAT applications in Medical Rehabilitation and potentially across an unlimited number of diagnostic groups. In

extending the participatory action research model and qualitative item development and review processes delineated by the PROMIS research team and refined by the Neuro-QOL project team, the authors have outlined a comprehensive step by step methodology for approaching qualitative rehabilitation data and developing targeted item banks that may be used for individuals with SCI, TBI, and other disabilities. Collecting and assimilating qualitative data in this manner serves to quantify the data, providing direct, measurable, and conceptually relevant information to researchers. Relative frequency of mention of each topic area is useful information when making the determination of which subdomains to include and which to omit. Given the inherently subject nature of the qualitative information elicited via participatory action research, and inability to account for subjective factors such as the presence of other individuals in a focus group setting, such qualitative information should not be mistaken for truly quantitative data and must therefore be considered alongside rather than instead of literature review and expert input. Combining grounded qualitative data with the rigors of item response theory and computerized adaptive testing yield an unprecedented ability to develop targeted, precise outcome measures for specific patient populations. For each domain, the end result is a precisely calibrated pool of items which are then carefully selected by CAT for a brief, sensitive, and specific measure that is appropriate for both research and clinical settings. The conceptual grounding of items serves to reduce or eliminate pitfalls of classically developed assessment instruments such as floor and ceiling effects. In the future, this methodology may be extended to other patient populations throughout the rehabilitation field and beyond to conceptually ground research on an individually-tailored, condition specific level. The authors are interested to see how other researchers may extend this methodology to address diverse research questions, including those unrelated to HRQOL.

Acknowledgments

This work was funded by the National Institutes of Health (NICHD/NCMRR and NINDS, 5R01HD054659) and the National Institute of Disability and Rehabilitation Research (H133N060022, H133G070138, H133A070037).

References

- Andresen EM & Meyers AR (2000). Health-related quality of life outcomes measures. *Arch.Phys.Med.Rehabil*, 81(12 Suppl. 2), S30–S45. [PubMed: 11128902]
- Bode RK, Lai JS, Cella D, & Heinemann AW (2003). Issues in the development of an item bank. *Arch.Phys.Med.Rehabil*, 84(4 Suppl. 2), S52–S60. [PubMed: 12692772]
- Brennan RJ (2006). Linking and equating. *Educational Measurement*, 6, 187–220.
- Carey JW, Morgan M, & Oxtoby MJ (1996). Inter-coder agreement in analysis of responses to open-ended interview questions: Examples from tuberculosis research. *Cultural Anthropology Methods*, 8(3), 1–5.
- Cella D (1995). Measuring Quality of Life in Palliative Care. *Semin Oncol*, 22 (2 Suppl 3): 73–81. [PubMed: 7537908]
- Cella D, and the Neuro-QOL Project Team. (2006). Neuro-QOL qualitative item review protocol Unpublished Manuscript.
- Cella D, Gershon R, Lai JS, & Choi S (2007). The future of outcomes measurement: Item banking, tailored short forms, and computerized adaptive assessment. *Quality of Life Research*, 16, 133–41. [PubMed: 17401637]
- Cella D, Lai J, & Item Banking Investigators (2004). CORE item banking program: Past, present and future, *Quality of Life*, 2, 5–8.

- Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. (2007). The Patient Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. *Medical Care*, 45(5 Suppl. 1), S3–S11.
- Charmaz K (1995). Grounded theory. In Smith J, Harre R & Van Langenhove L (Eds.), *Rethinking Methods in Psychology* (pp. 27–49). London, England: Sage.
- Choppin B (1968). Item bank using sample-free calibration. *Nature*, 219, 870–72. [PubMed: 5673356]
- Choppin B (1976). *Recent developments in item banking: Advances in psychological and education measurement* New York, NY: John Wiley.
- Choppin B (1978). *Item Banking and the monitoring of achievement* [Research in progress, Series No. 1]. Slough, England: National Foundation for Educational Research.
- Choppin B (1979). *Testing the questions: The Rasch model and item banking* Chicago, IL: University of Chicago, MESA Psychometric Laboratory.
- Choppin B (1981). Educational measurement and the item bank model. In Lacey C & Lawton D (Eds.) *Issues in evaluation and accountability* (pp. 87–113). London, England: Methuen.
- de Bruin AF (1997). The Sickness Impact Profile: SIP68, a short generic version. First evaluation of the reliability and reproducibility. *Journal of Clinical Epidemiology*, 50(5), 529–40. [PubMed: 9180645]
- DeWalt DA, Rothrock N, Yount S, & Stone AA, on behalf of the PROMIS Cooperative Group. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*, 45(5 Suppl. 1), S12–S21. [PubMed: 17443114]
- Division of Program Coordination, Planning and Strategic Initiatives. (2009). NIH roadmap for medical research Retrieved September 22, 2009, from <http://nihroadmap.nih.gov/>
- Flaugher R (1990). Item pools. In Wainer H (Ed.), *Computerized adaptive testing* (pp. 41–63). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Glaser BG, & Strauss AL (1967). *The discovery of grounded theory: Strategies for qualitative research* New Brunswick, NJ: Aldine Transaction.
- Gorden R (1992). *Basic interviewing skills* Itasca, IL: F. E. Peacock.
- Hall B Participatory research, popular knowledge, and power: a personal reflection. *Convergence* 1981;14:6–19.
- Hambleton RK (1989). Principles and selected applications of item response theory. In Linn RL (Ed.), *Educational Measurement* (3rd ed., pp. 147–200). New York, NY: Macmillan.
- Hays RD, Liu H, Spritzer K, & Cella D (2007). Item response theory analyses of physical functioning items in the medical outcomes study. *Medical Care*, 45(Suppl. 1), S32–S38. [PubMed: 17443117]
- Hays RD, Morales LS, & Reise SP (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care*, 9(Suppl. 2), II-28–II-42.
- Hill CD, Edwards MC, Thissen D, Langer MM, Wirth RJ, Burwinkle TM, et al. (2009). Practical issues in the application of item response theory: A demonstration using items from the Pediatric Quality of Life Inventory (PedsQL) 4.0 generic core scales. *Medical Care*, 45(Suppl. 1), S39–S47.
- Hoffman RR, Neville K, & Fowlkes J (2006). Using concept maps to integrate results from a cognitive task analysis of system development. In Canas AJ & Novak JD (Eds.), *Proceedings of the Second International Conference on Concept Mapping* San Jose, Costa Rica.
- Israel BA, Schulz AJ, Parker EA, & Becker AB (1998). Review of community-based research: Assessing partnership approaches to improve public health. *Annual Review of Public Health*, 19, 173–202.
- Kennedy P, Smithson E, McClelland M, Short D, Royle J, & Wilson C (2009). Life satisfaction, appraisals and functional outcomes in spinal cord-injured people living in the community. *Spinal Cord*
- Landis JR, & Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–74. [PubMed: 843571]
- Linacre JM, Heinemann AW, Wright BD, Granger CV, & Hamilton BB (1994). The structure and stability of the Functional Independence Measure. *Arch.Phys.Med.Rehabil* 75(2), 127–32. [PubMed: 8311667]

- MacQueen KM, McLellan E, Kay K, & Milstein B (1998). Codebook development for team-based qualitative analysis. *Journal of Cultural Anthropology Methods*, 10(2), 31–36.
- MetaMetrics. (1995). *The LEXILE framework for reading* Durham, NC: Author.
- National Spinal Cord Injury Statistical Center. (2007). *The 2007 annual statistical report for the Spinal Cord Injury Model Systems* Birmingham, AL: National Spinal Cord Injury Statistical Center.
- National Spinal Cord Injury Statistical Center. (2009). *National spinal cord injury database* Retrieved Oct. 1, 2009 from <http://www.spinalcord.uab.edu>
- Niemeier J, Kennedy R, McKinley W, & Cifu D (2004). The Loss Inventory: Preliminary reliability and validity data for a new measure of emotional and cognitive responses to disability. *Disability Rehabilitation*, 26(10), 614–23. [PubMed: 15204515]
- NIH Toolbox for the assessment of behavioral and neurological function (2009). Retrieved September 22, 2009 from <http://www.nihtoolbox.org>
- NVivo Qualitative Data Analysis Software (Version 8) [Computer Software] (2008). QSR International Pty Ltd.
- Quality of life in neurological disorders (2009). Retrieved September 22, 2009 from <http://www.neuroqol.org>
- Reeve B, Hays RD, Bjorner JB, Cook K, Crane P, Teresi J, et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care*, 45(Suppl. 1), S22–S31. [PubMed: 17443115]
- Revicki DA, & Cella DF (1997). Health status assessment for the twenty-first century: Item response theory, item banking and computer adaptive testing. *Quality of Life Research*, 6, 595–600. [PubMed: 9330558]
- Slavin MD, Kisala P, Jette A, & Tulsy DS (in press). Developing a contemporary outcome measure for spinal cord injury research. *Spinal Cord*
- Spector PE (1992). *Summated Rating Scale construction: An introduction* Newbury Park, CA: Sage Publications.
- Strauss A, & Corbin J (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Tourangeau R (1984). Cognitive sciences and survey methods. In Jabine TB (Ed.), *Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines* (pp.73–100). Washington, DC: National Academy Press.
- Tulsy DS, Victorson D, & Kisala PA (2010). Developing a contemporary patient reported outcomes measure for spinal cord injury research Unpublished Manuscript.
- Ware JE, Jr. (2004, Fall). Item banking and the improvement of health status measures. *Quality of Life*, 2, 2–5.
- Ware JE, Jr. & Sherbourne CD (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 30(6), 473–83.
- White GW, Suchowierska M, & Campbell M (2004). Developing and systematically implementing participatory action research. *Arch.Phys.Med.Rehabil*, 85(4 Suppl. 2), S3–12.
- Willis GB (1999) *Cognitive interviewing: A “how to” guide* Research Triangle Park, NC: Research Triangle Institute.

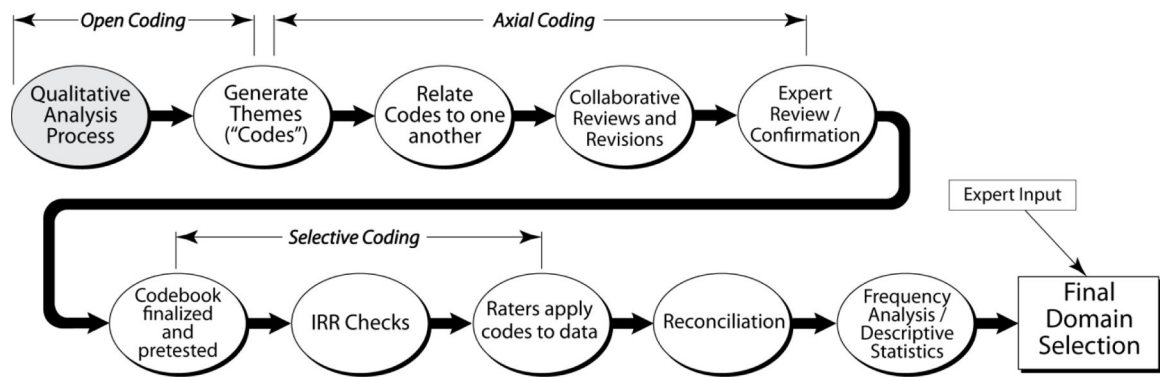


Figure 1.
Qualitative analysis process

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

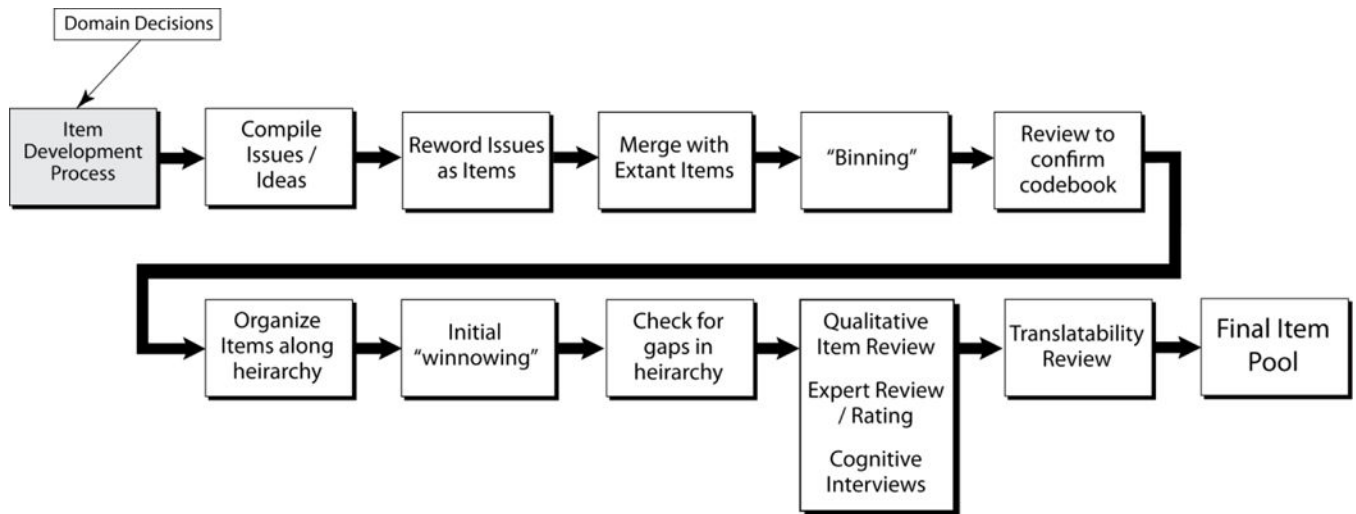


Figure 2.
Item development process

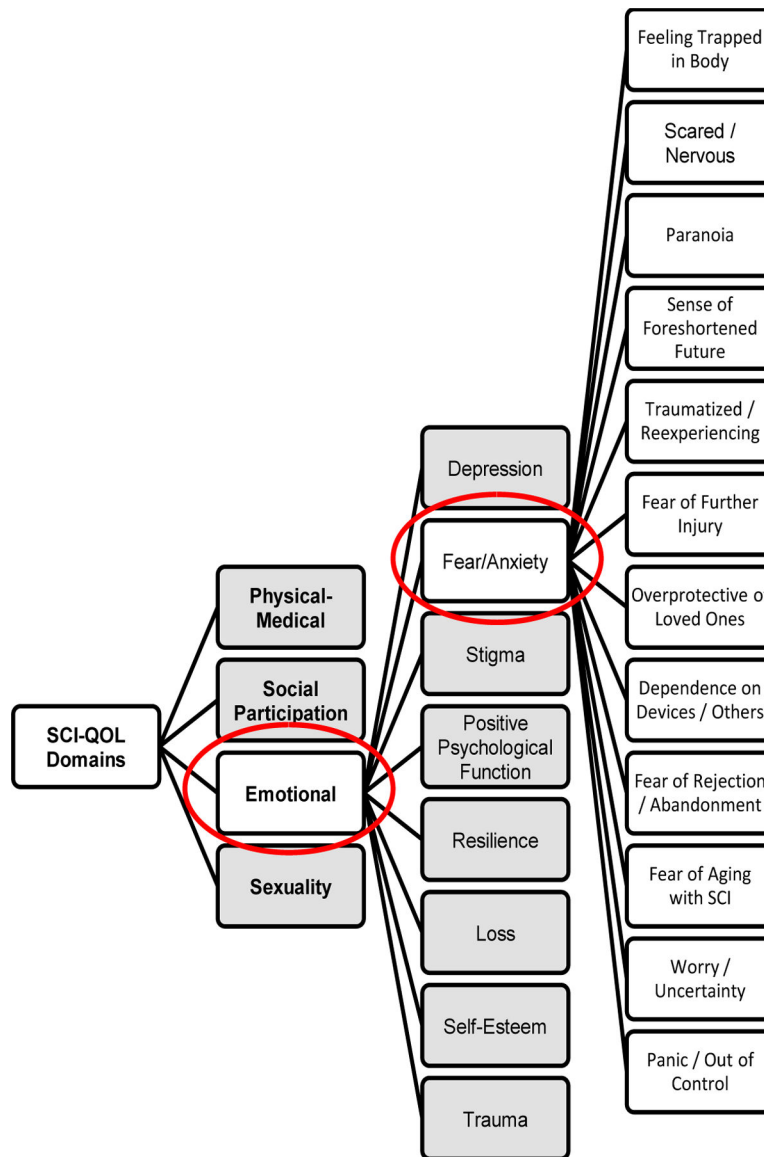


Figure 3.
Axial coding of the “Fear/Anxiety” subdomain

Table 1

Neuro-QOL Domain Structure

Neuro-QOL domains and subdomains
Physical health
Mobility
Upper Extremity / ADLs
Mental health - emotional
Depression
Anxiety / Fear
Positive Psychological Function
Stigma
Mental health - cognitive
Perceived Cognitive Function
Applied Cognitive Function
Social health
Social Role Performance
Social Role Satisfaction

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

SCI-QOL Domain Structure

SCI-QOL domains and subdomains	Overlap with Neuro-QOL?	Overlap with PROMIS?
Physical-functional health^a		
Mobility ^a	X	X
Upper Extremity / ADLs ^a	X	X
Physical-medical health		
Respiratory		
Skin / Pressure Ulcers		
Bowel		
Bladder		
Pain		X
Mental health - emotional		
Depression	X	X
Anxiety / Fear	X	X
Positive Psychological Function	X	
Stigma	X	
Grief / Loss		
Resilience		
Trauma		
Self-Esteem		
Social participation		
Social Role Performance	X	X
Social Role Satisfaction	X	X
Independence / Autonomy		
Sexuality		
Sexual Performance		
Sexual Satisfaction		

^aThe Physical-Functional Health domain was developed through a separate funding source (NIDRR grant #H133N060022 and #H133G070138) and is described in detail in Slavin, Kisala, Jette, & Tulsy (2009).

Table 3

Case Example. Issues Raised in Focus Groups and Corresponding Items

SCI-QOL: Patient focus group comment	Open Code	Axial Code	Selective Code	SCI-QOL item(s)
"It's hard enough to age getting old it's not fun...it's a lot of fear, like what's gonna happen down the line? How...am I gonna handle it or what's gonna happen to me. You get a lot of...anxiety, fear you know."	Fear / Anxiety	See Figure 3 for detail on the Axial Coding of the Fear/Anxiety Subdomain	<i>Fear of Aging with SCI</i>	I was afraid of what the future holds for me. ^a
"...(P)risoner in your own body..."; "...(F)eeling of like being trapped..."	Fear / Anxiety		<i>Feeling Trapped in Body</i>	I felt trapped in my own body. ^b
"I'm laying there and my chair is not next to where I am, I get anxiety."	Fear / Anxiety		<i>Dependence on Devices, Others</i>	I was anxious if my wheelchair / walking aid was not nearby. ^b
"If I fall and break a hip...then will it be the end of the road, will I not be able to have my own place, would it mean a nursing home?"	Fear / Anxiety		<i>Worry / Uncertainty</i>	I was afraid that I would be unable to live alone. ^b

^aNeuro-QOL item.New^bSCI-specific item.

Table 4

Case example: Frequency Data, Emotional Domain

Domain	Subdomain	Frequency of mention in SCI focus groups	Recommend for inclusion?
Emotional Health	Resilience	16%	Yes
	Grief / loss	14%	Yes
	Self-esteem	12%	Yes
	Depression	11%	Yes
	Positive psych. fxn	9%	Yes
	Anger / frustration	7%	No
	Independence / Autonomy	7%	Include in social: Independence / Autonomy
	Anxiety / fear	6%	Yes
	Stigma	3%	Yes

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript