Measuring Participation: The Patient-Reported Outcomes Measurement Information System Experience

Rita K. Bode, PhD, **Elizabeth A. Hahn, MA**, **Robert DeVellis, PhD**, and **David Cella, PhD on behalf of the PROMIS Social Domain Working Group**
Department of Physical Medicine and Rehabilitation, Feinberg School of Medicine (Bode), and the Medical Social Sciences (Hahn, Cella), Northwestern University, Chicago IL; Health Behavior and Health Education, University of North Carolina, Durham NC (DeVellis)

Abstract

Objectives—To describe the lessons learned in the initial development of PROMIS social function item banks.

Design—Development and testing of two item pools within a general population to create item banks that measure ability-to-participate and satisfaction-with-participation in social activities.

Setting—Administration via the Internet.

Participants—General population members (N=956) of a national polling organization registry participated; data for 768 and 778 participants used in the analysis.

Interventions—Not applicable.

Main Outcome Measures—Measures of ability-to-participate and satisfaction-with-participation in social activities.

Results—Fifty six items measuring the ability-to-participate were essentially unidimensional but did not fit an IRT model. As a result, item banks were not developed for these items. Of the 56 items measuring satisfaction-with-participation, 14 items measuring social roles and 12 items measuring discretionary activities were unidimensional and met IRT model assumptions. Two 7-item short forms were also developed.

Conclusions—Four lessons, mostly concerning item content, were learned in the development of banks measuring social function. These lessons led to item revisions that are being tested in subsequent studies.

Keywords

Quality of life; Rehabilitation

Corresponding author: Rita K. Bode, PhD, Dept of Physical Medicine and Rehabilitation, Feinberg School of Medicine, Northwestern University, 1823 Thatcher Ave, Elmwood Park, IL 60707, r-bode@comcast.net. Reprints will not be available from the author.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Social function (or social participation) is a common goal of persons who have been injured or debilitated by a chronic or acute medical condition. When asked their rehabilitation goals, rather than reduction of specific impairments or improvement in specific function, patients tend to report a more general desire to return to their previous level of activity or function. Social function therefore is an important secondary and long-term goal of rehabilitation—but one that is difficult to measure.

Development and validation of numerous item banks is the primary objective of the PROMIS cooperative group, a National Institutes of Health roadmap initiative[1]. In addition to developing item banks for pain, fatigue, physical function and emotional distress, PROMIS included the measurement of social health. To this end, a social health working group was formed to define a social health framework and construct pools of items to measure the most salient aspects of social health. As shown in the PROMIS Social Health framework (fig 1), social health is comprised of social function and social relationships, with social function further broken down into ability-to-participate and satisfaction-with-participation.

The term "social health" refers to a higher-order domain with measurable subdomains.[2] These sub-domains include social function (e.g., participation in society), social relationships (e.g., communication), and the quality, reciprocity and size of an individual's social network.[3, 4] Self-reported social health is broadly defined by PROMIS as perceived well-being regarding social activities and relationships. The two broad self-report outcomes under social health within the PROMIS framework are social function and social relationships.

Social function is defined by PROMIS as involvement in, and satisfaction with, one's usual social roles in life's situations and activities. These roles may exist in marital relationships, parental responsibilities, work responsibilities and social activities.[2, 5] Social function has also been referred to with terms such as role participation and social adjustment.[1] Qualitative and quantitative analysis of existing datasets and PROMIS data collected from 2005 to 2007 have resulted in a conceptual division of social function into "ability-to-participate" and "satisfaction-with-participation." Items were developed for activities with family, activities with friends, work activities and leisure activities.

This manuscript focuses on the development of the social function domains in PROMIS Wave 1. Following is a brief overview of the content development and analytic processes used in the development of these banks.

METHODS

Content Development

Development for all the PROMIS domains followed a series of steps beginning with the definition of the domains to be measured utilizing expert opinion and focus groups and progressing through a review of the literature to identify existing instruments and items to be included in the banks, writing new items, reviewing and revising new and existing items to better fit the domain definition, classifying and further reviewing the items using qualitative item review methods, and ending with the selection of items for Wave 1 testing. A complete description of the content development process is published elsewhere[6] and provided in the document titled "A Walk Through The First Four Years" on the PROMIS website[1]

The initial stage of content development involved the domain definition. The Delphi method was used with an expert panel to reach consensus on the domains to be measured. Two social function domains were proposed, ability-to-participate and satisfaction-with-

participation, each within three contexts: Family/Friends, Work/School, and Leisure activities. Focus groups were used to evaluate the fit of the proposed domain map with concepts identified by focus group participants as important aspects of social health. The results of this evaluation suggest that the conceptual model was comprehensive but needed to be further refined to more appropriately distinguish between responsibilities versus discretionary activities and situate the outcome of satisfaction as it relates to impairment in social and other domains of health.[7] Cognizant that the kind of social activities in which people engage may differ, it was decided to assess a person's ability-to-participate rather than actual level or frequency of participation. With this in mind, new items were written and existing items that measured frequency of performance were rewritten to ask about ability-to-participate.

An extensive literature search was conducted to identify existing instruments and items that measured the PROMIS social function domains. This search identified few relevant instruments and items (of the 660 items reviewed, only 6 were retained 'as is' and 28 were retained after revision), thereby necessitating the development of new items. Ninety-five new items were written and existing items revised to use a common rating scale in each subdomain: for ability-to-participate, a 5-point scale ranging from Never to Always and for satisfaction-with-participation, a 5-point rating scale ranging from Not at all to Very much were used.

Results of an earlier study with cancer survivors[8] suggested an ordering of items in four contexts, ranging from participation in leisure activities and social activities with friends (activities that were the first to be sacrificed) to work and family activities (activities that were the last to be sacrificed). New items were written to fill gaps within each of these contexts. When writing the new items to cover a wide range of ability-to-participate, variations in item types, such as including both positively-and negatively-worded items, synonyms for ability and satisfaction, and activity modifiers, were used (see below).

A previous study in which a physical function item bank was developed from existing sources[9] showed that, even when the responses were reverse coded, the most difficult items to endorse were positively worded. These results suggested that the inclusion of positively- and negatively-worded items might widen the range of the trait being measured. As a result, both positively- and negatively-worded items were written for the social function domains. The inclusion of both positively- and negatively-worded items was also intended to reduce the likelihood of response sets in answering items worded in the same direction. Psychological research on positive and negative affect has shown that directional opposites (e.g., *I am happy* versus *I am sad*) may not measure the same construct[10] but not much is known about directionally-worded opposites in the function domain (e.g., *I am able to …* versus *I am unable to …*). The PROMIS social function domain included directionally-worded opposites items.

Another variation in the item content was the use of synonyms within the positively- and negatively-worded items to reduce the potential of response set. For positively-worded ability items, *can do* was used interchangeably with *able to do;* for positively-worded satisfaction items, *satisfied with* was used interchangeably with *feel good about*. Similarly, for negatively-worded ability items, *unable to do* was used interchangeably with *limited in doing* and for negatively-worded satisfaction items, *dissatisfied with* was used interchangeably with *disappointed in* and *bothered by*.

Finally, to reflect different thresholds in ability and satisfaction, the content of some new items contained activity modifiers such as *that are important to me, that I want to do, that I expect I should do*, and *that people expect me to do*. Unmodified items, such as *I am able to*

perform my regular daily activities, were also included for comparison purposes. Thus we were able to determine how much easier or more difficult the item became when a modifier was used.

Wave 1 Testing Overview

The PROMIS protocol was approved by the institutional review boards at the participating research sites. Participants provided informed consent online.

Sampling and testing procedures. Items from 14 PROMIS domains were administered to a large general population sample (N=20,804) and a smaller clinical sample (N=1203) that was not used in this study. Two testing formats were used in this testing: one in which individual respondents were randomly assigned items from two related domains (full-bank testing) and another in which respondents were administered seven items from each domain (block testing). Both full-bank and block testing format was used with the general population but the clinical sample was administered items using just the block testing format. In both formats, items were administered online or by computer with each item displayed on a separate screen. A detailed description of the PROMIS Wave 1 testing is provided in the document titled "A Walk Through The First Four Years" on the PROMIS website[1]. Nine hundred and fifty six participants responded to items in the social function domains. Those who responded to fewer than 50% of the items, had response times of less than 1 second per item, or had repetitive responses to 10 consecutive items were excluded from the sample.

Analyses Conducted

The PROMIS analysis plan[11] consisted of a series of analyses with the goal of developing unidimensional sets of items that fit an IRT model and did not exhibit differential item functioning across gender, age, or education level.

Preliminary analyses—Data quality was assessed by examining the frequencies of responses to the items to identify any out-of-range responses, confirming the codes assigned for missing data, and assuring the rescoring of negatively-worded items. Preliminary analyses were conducted to identify unused or sparsely-used categories, examine whether the responses were monotonic, and examine the internal consistency of the responses to the items, using a criterion of Cronbach's alpha >.80. Items were deleted if categories were underused (less than 5 cases choosing the category) or if the corrected item-total score correlation was < .30.

Dimensionality analysis—For the dimensionality analysis, the Wave 1 sample was randomly split in half with each subsample used for either an EFA or CFA, both using Mplus version 4.1.[12] Dimensionality was initially assessed using CFA but given the number of untested items in the social domain pool, factor loadings were subsequently examined using EFA. For the EFA, polychoric correlations were analyzed using an unweighted least squares estimation procedure.[13] The number of factors was identified using eigenvalues greater than 1.0 and scree plots as criteria. Items loading .40 on a factor were examined to name the factor being measured. For the CFA, polychoric correlations were analyzed using a weighted least squares estimation procedure. Fit indices were used as evidence of unidimensionality with a minimum criterion of CFI>.90.[14] If fit to a 1-factor models was acceptable, the set of items was considered unidimensional. If fit was marginal, a bifactor model was used to determine if the data fit a model in which one general and several specific orthogonal factors could be assumed.[15] Acceptable fit using the bifactor model would identify the item set as 'essentially unidimensional'.[16] Finally, local dependence was

assessed by examining residual correlations among all item pairs, with correlations greater than .20 used as evidence of local dependence.

Parameter estimation and item response theory model fit—Parameters were estimated using the graded response model as implemented in Multilog.[17,18] Item characteristic curves were used to examine the distribution of responses across categories, item thresholds were used to examine the range of ability (or satisfaction) being measured by these items, slopes were examined to identify items with poor discrimination, and item information functions were used to estimate the range in which the most information was provided by an item. Likelihood-based chi-squared statistics (S-χ^2) were used to assess model fit.[19,20]

DIF analysis—For all unidimensional item sets, we examined uniform and nonuniform DIF using IRTLRDIF.[21] Uniform DIF detects differences across the entire theta range, whereas non-uniform DIF detects differences in only a segment of the theta range. We compared hierarchically nested IRT models; specifically, one model that fully constrained the parameters to be equal between two groups was compared to other models that allowed the parameters to be freely estimated. Three group comparisons were evaluated: by gender, by age (<65 versus 65) and by education (high school/General Educational Development or less versus higher education).

RESULTS

One hundred and eighty eight general population respondents taking the ability-to-participate and 188 taking the satisfaction-with-participation item sets were excluded from the sample. The remaining sample had an equal number of males and females and an average age of 51 (SD=17). They were predominately white (82%), with black and Hispanic respondents representing 7 and 10% of the sample, respectively. The most frequently reported clinical conditions were hypertension (36%), depression (23%) and arthritis (22%).

Preliminary analyses

The response distributions tended to be somewhat negatively skewed. Few category inversions were found and, where they occurred, they were at the bottom of the scale where the frequencies were small. The relatively sparse categories and inversions did not suggest a problem with respondent use of the rating scale. The alpha coefficients (.989 for ability-to-participate and .987 for satisfaction-with-participation) were high, but more informative was the generally acceptable (all >.40). the item-total correlations.

Dimensionality analyses

In the ability-to-participate EFA, 2 items loading on a separate "visiting relatives" factor and 5 that did not load on any factor were deleted. The remaining 49 items loaded on 3 factors: 21 discretionary activity; 16 social role and 12 discretionary activity limitations items. The items did not fit a 1-factor CFA (CFI=.843) but acceptable model fit (CFI=.912) was found when using the bifactor model. Better fit was found when treating 33 discretionary activities and 16 social role items separately, each with two specific factors. For discretionary activities, CFI=.945 and for social roles, CFI=.970. All the items loaded stronger on the general than specific factors and no local dependence was found. These two subdomains were considered 'essentially unidimensional,'

In the satisfaction-with-participation EFA, 30 items worded in terms of bother or disappointment were deleted. The remaining 26 items loaded on 2 factors: 12 discretionary activity items and 14 social role items. The full item set did not fit a 1-factor CFA (CFI=.

843) but acceptable model fit (CFI=.912) was found when using the bifactor model. Better fit was found when treating the discretionary activities and social role items separately in a bifactor model. For discretionary activities, CFI=.959 and for social roles, CFI=.968. All the items loaded stronger on the general than specific factors and no local dependence was found. These two subdomains were considered 'essentially unidimensional,'

Item response theory analyses

For the ability-to-participate items sets that were identified as essentially unidimensional, all categories were modal at some point in the scale but with sparse coverage at the top. The threshold range (−1.54 to 1.54) was narrower than the item measure range (−3.39 to 2.31); the slopes were acceptable for all items (1.84 to 4.52); and the theta estimates were most precise in the middle range of the trait. Despite being 'essentially unidimensional', 17 ability-to-participate items misfit the unidimensional IRT model when analyzed together and separately by specific factor. For the satisfaction-with-participation items sets that were identified as essentially unidimensional, all categories were modal at some point in the scale but with sparse coverage at the top. The threshold range (−1.21 to 1.15) was narrower than the item measure range (−2.68 to 2.02); the slopes were acceptable for all items (2.66 to 4.46); and the theta estimates were most precise in the middle range of the trait. Despite being 'essentially unidimensional', when all items were calibrated, 20 items misfit the unidimensional model. However, when calibrated separately for social role and discretionary activities, the subdomains fit the IRT model.

DIF analysis—For the two satisfaction-with-participation subdomains, no items exhibited nonuniform DIF and one (*satisfied with ability to do things for fun at home*) exhibited a trivial level of uniform age DIF (chi-square=5.64; p=0.02).

Two short forms were also developed (satisfaction-with-participation in social roles and satisfaction-with-participation in discretionary activities), each consisting of seven items.

DISCUSSION

Although the results of the psychometric analyses were not as good as we had hoped, we have learned some valuable lessons and have the beginnings of two item sets that can be developed into banks in future studies. The social function domain was the least developed and tested of all the PROMIS domains[22] but, through this work, definitions of these domains have been refined. We are optimistic that the ordering of items from those involving home and family activities to those involving activities outside the home and involving a larger social circle suggested by these results will be borne out with additional research. We further believe that the distinction between items measuring social roles and discretionary activities is conceptually meaningful and can lead to more precise measurement of participation in social activities. In addition, we learned a number of valuable lessons regarding the content of social function items.

Lesson 1: Combining positively- and Negatively-Worded Items

Although the use of both positively- and negatively-worded items may widen the range of the traits that an instrument could reliably measure and reduce response set, combining these item types in this study threatened the dimensionality of the combined item set. After the initial CFA showed unacceptable dimensionality, exploratory analyses were conducted to examine the factor structure of the item sets. In both the ability-to-participate and satisfaction-with-participation domains, positively- and negatively-worded items loaded on different factors. The bifactor model was used to determine if the two item types were secondary factors under a general factor, that is, whether the items were 'essentially

unidimensional.' However, even with the assumption of secondary factors, the item sets were not undimensional. Closer examination of the items loading on each factor showed the relevant distinction to be whether the item was measuring ability-to-participate in social roles (activities with family and work activities) or ability-to-participate in discretionary activities (activities with friends and leisure activities). The positive-negative split in the ability-to-participate items appeared to be less pronounced in the social role than in the discretionary activity items, suggesting an as-yet-unknown distinction in the perception of ability-to-participate in social activities with others. Negatively-worded satisfaction-with-participation items had been deleted from the item pool due to lack of dimensionality thus it was not possible to determine whether there was a similar distinction in these items. Because data from the clinical sample were not used in the analysis of social function items, it is not clear if the problem with combining positive and negative items was related to using a general population instead of a clinical sample or if the two sets of items are measuring slightly different constructs.

Lesson 2: Effect of Using Non-Synonymous Terminology

While the intent was to choose synonyms in the wording of the ability-to-participate and satisfaction-with-participation items, in some items the wording turned out to be a problem. For ability, there seemed to be little effect from using synonyms in the wording of the items. Items with similar content had approximately the same likelihood of endorsement (e.g., the theta estimates for *feel good about ability to do things for family* and *satisfied with ability to do things for family* were .28 and .30, respectively), did not differ in their discrimination ability, and loaded on the same factor. However, for the satisfaction-with-participation items, the use of *disappointed with* and *bothered by* proved to be problematic. Indications of this problem were found in the relationship of responses to these items and a set of global items (one for each of the PROMIS domains) that were administered along with the domain item sets. The satisfaction-with-participation items using these synonyms not only loaded on a different factor than the positively-worded items but their correlations with the scores on the mental health global item *(rate your mental health from excellent to poor)* were as high or higher than with the scores on the social activity global item (*rate your social activity/ relationships from excellent to poor*). Spearman rho correlations of these satisfaction items with the global social activity item ranged from .34 to .54 and with the mental health global item ranged from −.32 to −.51. As a result of this ambiguity, the 29 negatively-worded satisfaction items were deleted from the item sets, reducing content coverage and leaving fewer items for examination.

Lesson 3: Effect of Wording Modifications

Modifiers were used with the ability items to reduce ceiling effects, but too little effect. The correlations among four versions of an item asking about ability-to-participate in leisure activities with friends ranged from .79 to .83. Asking about ability-to-participate in leisure activities or those with friends showed a similar pattern in the likelihood of endorsing the item. For the social function domains, likelihood of endorsement refers to the extent to which respondents are more or less able to participate (or satisfied with their participation) in various activities. Activities that are less likely to be endorsed are those in which the sample reported less ability (or satisfaction) and activities that are more likely to be endorsed items are those in which the sample reported more ability (or satisfaction). In both cases using the (*activity) that is important to me* modifier tended to make the item slightly more likely to be endorsed and using the (*activity) that I want to do* modifier tended to make the item slightly less likely to be endorsed but not substantially so. The likelihood of endorsement when framing items in terms of one's own or others' expectations were similar to the likelihood of endorsement when no modifiers were used. The pattern was different for items asking about ability-to-participate in work activities. In these items, ability to meet

people's expectations were the most likely to be endorsed and ability to meet one's own expectations were least likely. A problem with using these modifications is they confound the construct definition such that harder-to-endorse items are measuring a slightly different construct (one reflecting personal expectations) than easier-to-endorse items (one reflecting societal expectations).

Lesson 4: Confirming an a Priori Item Ordering

In a previous study in which similar items were administered to a sample of cancer patients [2], discretionary activities and social roles loaded on a single factor, However, in this study using data from this general population in which the majority of the participants reported some common chronic conditions and few reported more serious conditions such as stroke or spinal cord injury, these items did not load on a single factor. The difference in results could be attributed not only to the type of sample (clinical versus general population) but perhaps to the IRT model used (Rasch analysis in the original study and 2pl in this study). The results from this study, however, suggest a partial confirmation of the hypothesized ordering of items in terms of likelihood of endorsement in that social role items (participation in work and family activities) are more likely to be endorsed than the discretionary activities items (participation in leisure activities and activities with friends). That is, individuals tended to receive higher social role scores than discretionary activities scores. Furthermore, within the ability and satisfaction subdomains, the previously suggested hierarchy was also suggested. Items measuring satisfaction-with-participation in social roles involving family or household activities were more likely to be endorsed than those involving work activities. Similarly, items measuring participation in discretionary activities involving activities around the home were more likely to be endorsed than those involving activities outside the home. In both the social role and discretionary activities subdomains, there was a similar progression of activities; that is, activities in which people were more able to participate or be satisfied with their participation tended to be those that could be performed at home or with one's family and activities in which people were less able to participate or be satisfied with their participation were those performed outside of the home or with a larger social network.

Lesson 5: Unidimensionality Versus Item Response Theory Model Fit

While not an issue in content development, discrepancies between dimensionality and model fit may be related to the consistency of the ordering of items by likelihood of endorsement across individuals. Items at the bottom of the scale (easier items to endorse) were consistently endorsed by most individuals while those at the top of the scale (harder items to endorse) were consistently endorsed by few individuals; however, the endorsement of items in the middle of the scale may be less consistent. Thus, despite all items essentially measuring the same construct, this lack of a common ordering of items across respondents can cause a problem. This suggests that a consistent item ordering across individuals is as important as unidimensionality in developing an item set that has acceptable IRT model fit. In the social function domains, item sets that were essentially unidimensional using the bifactor model did not necessarily have acceptable IRT model fit. In the satisfaction-with-participation subdomains that had acceptable model fit, the hierarchy was more evident than in the ability-to-participate subdomains. Perhaps this is an issue with the model assumptions in that a unidimensional IRT model may only work well with items that strictly fit a 1-factor CFA model (that is, a single general factor) whereas a multidimensional IRT model may be needed with items identified as only essentially unidimensional, that is, reflecting specific secondary factors in addition to the general factor). Perhaps the IRT model fit assumptions are too stringent for a small number of items and a large sample. Or perhaps preferences in the choices people make regarding the activities in which they participate preclude the

development of a hierarchy of items measuring participation in social activities. Further research is required to examine these possibilities.

Insights and Future Directions

The PROMIS experience with measuring social function had mixed results. Little foundation existed on which to build these banks, necessitating the development of many new items. Although the items developed for the two original social function domains failed to meet PROMIS model assumptions, much has been learned that should help in future development.

The optimal level of social function may not be determinable a priori. While more satisfaction is considered better, more activity may not necessarily be better. Perhaps a narrower definition of the domains is needed to isolate aspects of participation that are hierarchical in nature. Or perhaps a certain threshold of limitation needs to be exceeded before one curtails participation in social activities. Satisfaction-with-participation may be more relevant than ability-to-participate. Regardless of level of activity; being more satisfied may be a better indicator of someone's social function than the level of their activity. The fact that both satisfaction-with-participation subdomains fit the IRT model may suggest this is the case.

Personal preference plays a large part in the extent to which people engage in different social activities and additional research is needed to distinguish what people want to do from what they actually do. Perhaps discrepancies between wanting and doing can be used to better understand the role that activity preference plays in measuring activity participation and result in hypotheses regarding the ordering of activities into a generalizable and meaningful hierarchy.

It may be that differences in which activities people participate cancel out when summarized at the group level. Other than identifying activities endorsed by most people and those endorsed by very few people, the levels of endorsement of activities between these extremes may be too similar to be hierarchical. The hierarchies may differ for as-yet-unidentified subsets of people and would need to be further explored.

Social function in a general population may differ in substantive ways from social function in a clinical sample. Physical limitations may not result in reduced participation until the limitation is more severe. Additional testing and development with a more heterogeneous sample may detect a currently hidden hierarchy.

Two type of participation were identified in both social function domains that are separate but related: participation in social roles (what you do for others) and participation in discretionary activities (what you do with others). Although the constructs being measured were distinct, social role items are more likely to be endorsed than discretionary activity items.

The sizes of the current item sets are insufficient for banking purposes and more work is needed to develop and refine items that measure both types of social function. Sets of 12 and 14 satisfaction-with-participation items can be considered a basis upon which to develop banks that are of adequate depth for this trait across the continuum. However, different ways of measuring ability-to-participate may be needed. Explorations of alternative item types, such as asking about interference or building attributions to the health condition into the items, should be considered in developing such banks.

Immediate Steps

Work is already under way to apply what we learned in the first wave of PROMIS testing. Items have been rewritten so that the subdomains shown in figure 1 contains only positively- or negatively-worded items; disappointment and bother are no longer used in the wording of satisfaction-with-participation items, and modifiers regarding expectation have been eliminated. The new items are being administered to persons with cancer, arthritis and stroke as part of a supplemental PROMIS study. Analysis of the data from these administrations will provide a test of the insights we've gained and should go a long way in improving our measures of social function.

Acknowledgments

List of Abbreviations

CFI	Comparative Fit Index
DIF	differential item functioning
CFA	confirmatory factor analysis
EFA	exploratory factor analysis
IRT	item response theory
PROMIS	Patient-Reported Outcomes Measurement Information System

References

1. [Accessed November 4, 2009.] PROMIS: A Walk Through the First Four Years. Available at: http://www.nih.promis.org

2. McDowell, I. Measuring health: a guide to rating scales and questionnaires. Oxford: Oxford Univ Pr; 2006.

3. Beels CC, Gutwirth L, Berkeley J, et al. Measurements of social support in schizophrenia. Schizophrenia Bulletin. 1984; 10:399–411. [PubMed: 6382589]

4. Brekke JS, Long JD, Kay DD. The structure and invariance of a model of social functioning in schizophrenia. J Nerv Ment Dis. 2002; 190:63–72. [PubMed: 11889358]

5. Dijkers MP, Whiteneck G, El Jaroudi R. Measures of social outcomes in disability research. Arch Phys Med Rehabil. 2000; 81(12 Suppl 2):S63–80. [PubMed: 11128906]

6. DeWalt DA, Rothrock N, Young S, Stone AA. on behalf of the PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. Medical Care. 2007; 45(5 Suppl 1):S12–21. [PubMed: 17443114]

7. Castel LD, Williams KA, Bosworth HB, et al. Content validity in the PROMIS social-health domain: a qualitative analysis of focus-group data. Quality of Life Research. 2008; 17:737–49. [PubMed: 18478368]

8. Hahn EA, Cella D, Bode RK, Hanrahan RT. Measuring social well-being in people with chronic illness. Social Indicators Research. 2009 Epub ahead of print.

9. Bode RK, Cella D, Lai J-S, Heinemann AW. Developing an initial physical function item bank from existing sources. J Applied Measurement. 2003; 4:124–36.

10. Schmukle SC, Egloff B, Burns LR. The relationship between positive and negative affect in the Positive and Negative Affect Schedule. Journal of Research in Personality. 2002; 36:463–75.

11. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care. 2007; 45 (5 Suppl 1):S22–S31. [PubMed: 17443115]

12. Muthen, LK.; Muthen, BO. Mplus User's Guide. Los Angeles: Muthen & Muthen; 2006.

13. Muthen, BO.; du Toit, SH.; Spisic, D. [Accessed January 8, 2007.] Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. 1987. Available at http://www.gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf

14. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6:1–55.

15. Gibbons R, Hedeker D. Full-information item bi-factor analysis. Psychometrika. 1992; 57:423–36.

16. McDonald RP. The dimensionality of test and items. Br J Math Stat Psychol. 1981; 34:100–17.

17. Samejima F. Estimation of ability using a response pattern of graded scores. Psychometrika Monograph Supplement, No 17. 1969

18. Thissen, D. Multiple, categorical item analysis and test scoring using item response theory. Lincolnwood: Scientific Software International Inc; 1991. MULTILOG user's guide.

19. Orlando M, Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. Applied Psychological Measurement. 2000; 24:50–64.

20. Orlando M, Thissen D. Further examination of the performance of S-X^2, an item fit index for dichotomous item response theory models. Applied Psychological Measurement. 2003; 27:289–98.

21. Thissen D. IRTLRDIF -Software for the computation of the statistics involved in item response theory likelihood-ratio test for differential item functioning (Version 2.0b). 2003

22. Hahn EA, Cella D, Bode R, Rosenbloom S, Taft R. Social well-being: The forgotten health status measure. Quality of Life Research. 2005; 14:1991. (abstract #1576).
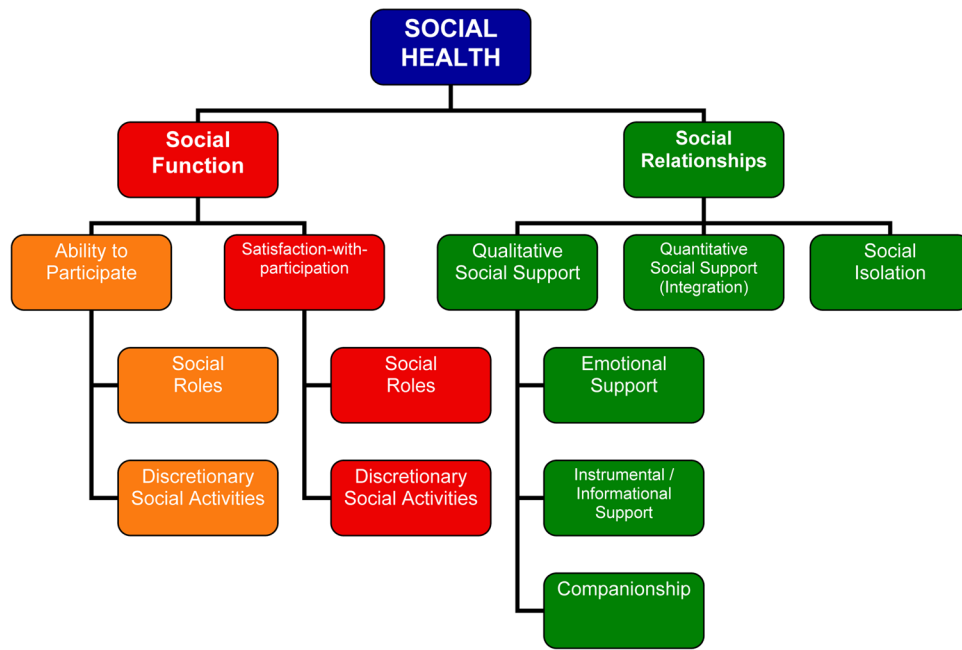
Fig. 1.
PROMIS Social Domain Framework.
© Copyright 2008 by the PROMIS Cooperative Group and the PROMIS Health
Organization Reprinted with permission.