Assessing the factor structure of a role functioning item bank

Milena D. Anatchkova,
Department of Quantitative Health Sciences, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01665-002, USA

John E. Ware Jr., and
Department of Quantitative Health Sciences, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01665-002, USA

Jakob B. Bjorner
National Research Center for the Working Environment, Copenhagen, Denmark. i3 QualityMetric, Lincoln, RI, USA

Milena D. Anatchkova: Milena.Anatchkova@umassmed.edu

Abstract

Purpose—Role functioning (RF) is an important part of health-related quality of life, but is hard to measure due to the wide definition of roles and fluctuations in role participation. This study aims to explore the dimensionality of a newly developed item bank assessing the impact of health on RF.

Methods—A battery of measures with skip patterns including the new RF bank was completed by 2,500 participants answering only questions on social roles relevant to them. Confirmatory factor analyses were conducted for the participants answering items from all conceptual domains ($N = 1193$). Conceptually based dimensionality and method effects reflecting positively and negatively worded items were explored in a series of models.

Results—A bi-factor model (CFI = .93, RMSEA = .08) with one general and four conceptual factors (social, family, occupation, generic) was retained. Positively worded items were excluded from the final solution due to misfit. While a single factor model with methods factors had a poor fit (CFI = .88, RMSEA = .13), high loadings on the general factor in the bi-factor model suggest that the RF bank is sufficiently unidimensional for IRT analysis.

Conclusions—The bank demonstrated sufficient unidimensionality for IRT-based calibration of all the items on a common metric and development of a computerized adaptive test.

Keywords

Role functioning; Bi-factor model; Dimensionality

Introduction

The World Health Organization defines health as not only the absence of disease but also the presence of physical, mental, and social well-being [1]. Role functioning (RF) is a key component of social well-being and thus an important outcome in health research. Within this framework, social roles can be viewed as the building blocks for an individual's societal

Correspondence to: Milena D. Anatchkova, `Milena.Anatchkova@umassmed.edu`.

involvement as a player in socially recognized areas of human effort [2], and the importance of the impact of poor health on this area cannot be overstated.

Role impairment and role disability are associated with a large number of chronic and mental health conditions. A recent study of the impact of mental and physical conditions on role disability using nationally representative data revealed that 53% of US adults have one or more of the 30 studied conditions [3]. Afflicted respondents reported on average 32.1 more disability days in a year compared to healthy matched controls, which was translated to 3.6 billion days of role disability days in the population [3]. Both mental and chronic medical disorders lead to significant role impairment but on different aspects of RF [4]. These results have led to the increased recognition of role disability as a major source of indirect cost of illness and disease burden [3], making it an important construct to assess accurately.

A number of tools measuring the severity of role impairment and functioning have been proposed. Existing measures are quite diverse in their theoretical approach, format, focus of assessment, comprehensiveness, psychometric properties, areas of use, and popularity with researchers [5]. In our recent work, we have been using these tools to develop a generic role functioning item bank using a multistage process, which is described briefly in the methods section. We have already reported the results of the qualitative stages of this work, resulting in the formulation of 75 generic items assessing health impact on functioning across 3 role areas(family life, occupational life and social life) [5]. Some of the developed items used positive wording aiming to broaden the measurement range of the bank (see Table 2 for item text).

The main focus of this paper is to describe the results of the next steps in the process of item bank development, namely the exploration of the factorial structure of the newly developed item bank and the evaluation of the assumption of sufficient unidimensionality for item response theory (IRT) model estimation. IRT models allow the selection of the most appropriate items for each respondent through computerized adaptive testing (CAT) [6–9] of RF. This is a promising new strategy in the development of improved health outcome measures [10, 11].

Many approaches have been employed in the evaluation of dimensionality [12, 13]. Both exploratory and confirmatory factor analyses (EFA and CFA) have often been used for that purpose, with CFA having the advantage of allowing hypothesis testing [13]. Supplemental strategies in the assessment of dimensionality are bi-factor analysis [14], parallel analyses [15], comparisons between IRT slope parameters computed from larger and smaller sets of the items [16] and tests of methods effects based on multi-trait multi-method analyses [17–19].

We have favored an investigative approach in our work in order to build a model that is conceptually meaningful, while providing sufficient information regarding the measurement requirement of unidimensionality. We used a series of CFA that include bi-factor models in our exploration of the bank's dimensionality. We hypothesized that both item content and item wording can have an effect on the item bank's factorial structure. The goal of this study was to determine whether these effects would allow the role functioning bank to be considered sufficiently unidimensional for IRT applications. Details about the tested models are provided in the methods section.

Methods

Participants

A total of 2,500 participants were recruited for the study via the Internet by a panel company (www.YouGovPolimetrix.com) in the winter of 2008. To ensure higher variability of responses and better generalizability of results, we aimed to recruit a sample that was equally stratified across age groups (18–34, 35–54, 55+) with gender, race and ethnicity quotas representing the US population. A little more than half of the sample ($n = 1,445$) included participants with selected chronic conditions and the other half was recruited as general population, but excluding participants with these conditions ($n = 1,055$). As a result of the skip pattern, there were items with systematically missing data. The models exploring the dimensionality of assessment were developed with the data from the subsample (SS1) of participants who responded to questions from all conceptual domains but excluded items on parent and partner role ($N = 1193$). Several of the best fitting models were tested in a second subsample (SS2) of participants ($N = 414$) who were living with a partner and taking care of a child and thus answered additional questions on the parent and partner role as well.

All participants completed a consent form and received an incentive for their participation in the study in the form of "polling points".

Instruments

The bank included newly developed items assessing role functioning, items from an existing short role functioning bank and 4 items from the Role Physical Scale of the SF-36 Health Survey [20]. All items were presented consecutively one item per screen.

Item development process—A multi-stage process of item bank development is required for the implementation of a computerized test. The details, challenges, and analytic decision required in the process have been described in detail in earlier publications [10, 16, 21]. Briefly, the process goes through the following steps: 1/Construct definition—the construct is defined based on theory and previous empirical results; 2/Item development—review of existing measures and formulation of new items; 3/Data collection for item calibration and testing; 3/Evaluation of assumptions of unidimensionality and local independence; 4/Psychometric analyses and test of model fit; 5/CAT simulations and specifications of a final CAT. We have followed this process in the development of a generic RF item bank.

Conceptual model—Our theoretical conceptualization of the role functioning construct was inspired by the biopsychosocial model of health and disability and the International Classification of Functioning, Disability and Health (ICF) [1, 5]. The conceptual model defined role functioning as involvement in life situations related to family life, partner relationship, household chores, work for pay, studies, social life (including interactions with friends), leisure time activities, community involvement (including volunteer work), and everyday living activities. Consistent with the ICF, the influences of personal and environmental factors on role functioning were postulated, focusing specifically on the effects of life stage, choice, and opportunity for participation in specific roles [5].

We used review of the literature, including review of existing measures and the results of a focus group study in the formulation of the measurement model and item bank items [5]. The measurement model for the construct of health impact on role functioning covers three interrelated domains of social, occupational and family roles and functioning. The social domain includes all role activities related to the community, leisure, social life and entertainment. The occupational role domain includes employment roles and activities as

well as the student role. The family domain includes family roles, including the role of a partner as well as household chores and activities.

New items—Seventy-five items assessing the impact of health on role functioning across the model domains (family life ($k = 29$, 4 positive), occupational life ($k = 28$, 2 positive), social life ($k = 18$, 3 positive) were developed. Each item was designed to cover only one of the content domains. All of the items included health attribution (e.g. "health" was specified as the factor influencing the level of functioning) and a 4-week recall period. Most of the items were worded with negative attribution to health e.g. "Did your health make it hard to start work on time over the past 4 weeks?", but some positively worded items were also included e.g. "Did your health allow you to perform at your best during the past 4 weeks?" All of the new items were presented one per screen. Skip patterns were used for items assessing role functioning in areas that were not relevant for participants (e.g. unemployed people were not asked questions on occupation).

Short RF bank [22]—Eight items were included from a previously developed item bank on role functioning. The items included attributions to health and a 4 week recall. Each of the items covered more than one of the identified content domains e.g. "In the past 4 weeks, how much of the time have you had difficulty in performing work or other daily activities, because of your health?"

Role physical scale items—Four items of the Role Physical Scale of the SF-36 Health Survey were also included in the bank.

Statistical analyses

We used an investigative approach [13] to the assessment of dimensionality including eigenvalue analyses and a series of confirmatory factor analyses (CFA) models to assess the factorial structure of the item bank and the potential presence of wording method effects. Three types of CFA models were hypothesized and subsequently tested. First, we tested models that included only conceptual factors (Models 1a–1d). As a second step, we tested a number of models including both conceptual factors and methods factors aiming to test the presence of wording effects in the items (Models 2a–2c), following an approach that has been previously used in a number of studies on various content scales [19, 23]. As a final step, we tested bi-factor solutions (Models 3a–3c), following previous recommendations to use this approach in the assessment of sufficient unidimensionality of item banks with theoretically driven subdomains [14]. More specifically, we tested the following model types:

Health impact concepts models (Models 1a–1d)—We hypothesized four different models representing only substantive factors of role functioning (Fig. 1). Model 1a represents a single (unidimensional) factor model of health impact on role functioning. Model 1b is a hierarchical model representing three content domains of role functioning identified as important in previous studies, namely family life, occupational life, and social life [5]. Model 1c postulates one additional "generic" domain, since the bank included items from the Short Role Functioning Bank and SF-36 Role Physical items that were covering more than one of the three content areas described above. Models 1a–1c are nested in models 2a–2c below. Model 1d is a two factor model with two separate substantive factors for positive and negative health impact corresponding to positively and negatively worded items.

Health impact concepts and methods components models (Models 2a–2c)—Correlated traits, correlated methods framework was used for these models [23] specified by

adding two wording method factors for the one, three and four health impact concepts models described above (Fig. 2). All items loaded both on a content factor and on either positive or negative methods factor depending on their wording. All method factor loadings were constrained to be equal, and correlations between method and conceptual factors were fixed to zero.

Bi-factor models (Models 3a–3c)—We also used a series of bi-factor models to evaluate unidimensionality and the effect of positive wording of items (Fig. 3). We planned to test a bi-factor model with one general impact factor and four subdomain group factors as described above in the four factor hierarchical model (Model 3a, 3c). All factor item loadings in the models were freely estimated, while factor variances were fixed to identify the model. Factors were assumed to be uncorrelated. We also planned to combine the two approaches and test a bi-factor model that is also taking into account the difference in the item wording through additional methods factors. Model 3a is nested in Model 3b, which strictly speaking is not a bi-factor model.

Models including items on partner and children (Models 4a–4d)—We planned to test a selected number of models with SS2 based on the results of the analyses completed with SS1. As results strongly suggested the presence of wording methods effects, we included models that reflected these results. Conceptually models 4a–4d in SS2 correspond to models 2a–2c, and 3c, but as stated earlier they included some additional items in the family domain. All analyses were conducted using MPlus (3.11) software for analyses of categorical data and polychoric correlations with weighted least-squares with mean and variance adjustment (WLSMV) parameter estimation [24]. Despite the wide use of CFA in dimensionality testing, the results of a recent study have suggested that published criteria and cut-off scores for CFA fit indexes are influenced by the number of items and the non-normality of the data and thus may not be appropriate when used as a single approach in the evaluation of dimensionality of large item banks [13]. Thus, patterns of standardized item loadings were examined and used in the model evaluation along with two widely used fit indexes—the comparative fit index CFI [25] and the root-mean-square error of approximation [25] with values of CFI >.90 and RMSEA <.10 for acceptable model fit [26, 27]. Finally, residual correlations were examined to evaluate local misfit. Once the best fitting model was identified, the question of sufficient unidimensionality was evaluated based on the model parameters, in particular the item loadings [12, 14, 28, 29].

Results

Sample characteristics

The mean age of respondents in the entire sample of 2,500 participants was 49 years (range 18–94), 51% were females, 76% Caucasian, 19% of the sample had an educational level lower than high school, 34% were college graduates, and 17% had post-graduate degrees. Proportions of people with chronic conditions were asthma $n = 416$ (17%), heart disease $n = 333$ (13%), diabetes $n = 367$ (15%), auto-immune diseases $n = 329$ (13%). Participants in SS1 had slightly lower mean age of 47 years (range 18–94) but were comparable on other demographic characteristics. Proportions of people with chronic conditions in this subsample were asthma $n = 212$ (18%), heart disease $n = 116$ (10%), diabetes $n = 173$ (15%), auto-immune diseases $n = 107$ (9%). Participants in SS2 were younger (mean age 44), predominantly male (55%) and had higher proportion of people with asthma $n = 97$ (23%) and lower proportions of people with heart disease 31 (8%) and auto-immune conditions $n = 28$ (7%).

Item characteristics

We defined high scores to indicate higher level of functioning, and all items (rating scale 1–5) were scored in that direction. Descriptive statistics showed skewed distribution for the data (mean = 4.20, standard deviation = 1.08, skewness = −1.48, kurtosis = 1.70). Due to the skewed responses, some of the items had very low frequency count cell for some of the response options, so we collapsed response categories for 11 of the items before proceeding with the analyses.

Factor analyses results by model—In eigenvalue analyses, 4 components had eigenvalues greater than one (53.2, 2.6, 1.9 and 1.7), and the large ratio of the first value to the second suggested unidimensional structure of the bank. The results of the hypothesized CFA's are presented below, and goodness of fit indexes for all CFA models are presented in Table 1.

Health impact concepts models (Models 1a–1d)

Model 1a: The fit of the one-factor concept model was poor (CFI = .840, RMSEA = .158). Standardized factor loadings ranged from .680 (item soc15) to .946 (item RB4, RB5).

Model 1b: The hierarchical model with three content domains, where positive items were not modeled, did not improve model fit (CFI = .861, RMSEA = .134). The three domains had item loadings in ranges from .755 (item fam4) to .986 (item RB2) in family life, from .709 (item soc15) to .962 (item soc12) in social life, and from .714 (item work29) to .960 (item RB4, RB5) in occupational life. Second order factor loadings ranged from .935 to .977.

Model 1c: The fit of the four factor hierarchical model was similar to the two factor conceptual model (CFI = .872, RMSEA = .122). Item loadings ranged from .773 to .926 for family life, .727 to .941 for occupational life, .711 to .960 for social life and .884 to .967 for the generic domain. The second-order factor loadings were high (.929–.970).

Model 1d: Modeling positive items as a separate factor of positive health slightly improved fit indexes, but the overall fit of the model was still unacceptable (CFI = .883, RMSEA = .132). Standardized factor loading ranged from 0.847 (item soc15) to .938 (item work28) in the positive health factor, and from .729 (item Fam4) to .947(item RB4, RB5) in the negative health factor. The correlation between the two factors was .721.

Health impact concepts and methods components models (Models 2a–2c)

Model 2a: The addition of methods effects to the conceptual models improved model fit. However, the one concept model with 2 wording effects still had a relatively poor fit (CFI = .884, RMSEA = .131), item loadings range .476–.890.

Model 2b: The model with three conceptual factors and 2 wording methods had a marginally acceptable fit (CFI = .908, RMSEA = .107), with item loading ranges of .633–.893 in family life, .720–.866 in social life, and from .642 to .865 in occupational life. The range of second-order factor loadings was .927–970.

Model 2c: The hierarchical model with 4 conceptual factors and two methods effect had an acceptable fit (CFI = .925, RMSEA = .09). The item loadings for the four factors, respectively, were for family life from .741 to .898, for social life from .506 to .932, for occupational life from .680 to .915 and for the generic impact factor from .470 to .968. The range of factor loadings on the impact factor was .914–.965.

Bi-factor models (Models 3a–3c)

Model 3a: The bi-factor model with 4 group factors had a marginally acceptable fit (CFI = .895, RMSEA = .108). The item loadings for the overall health impact factor were in general higher (item loadings range .588–.946) than the item loadings in the four group factors which, respectively, were for family life from .217 to .576, for social life from .116 to .696, for occupational life from −.023 to .464 and for the generic impact factor from .084 to .437. The highest loadings in the group factors were for the positively worded items.

Model 3b: For this model, we added a wording methods effect to Model 3a. In this stage we decided to model only the positively worded items because they had higher items loadings in Model 2c. In addition, residual correlations higher than .2 were present in Model 3a only among positively worded items and they were the items with highest loadings on group factors. The bi-factor model with methods effect factor for positively worded items had acceptable fit (CFI = .933, RMSEA = .084). The item loadings for the overall health impact factor were higher (item loadings range .587–.953) than the item loadings in the four group factors which, respectively, were for family life from .208 to .526, for social life from .200 to .403, for occupational life from .016 to .424, for the generic factor from .044 to .430. The item loadings for the positively worded items were high (.599).

Model 3c: The bi-factor model with positive wording factor fit the data best, but the high factor loadings on the methods factor suggested problems with the fit of these items in the underlying unidimensional construct. In addition, comparison of item loadings in the unidimensional model and the general impact factor of the bi-factor solution revealed some large differences, suggesting that positively worded items are deviating from the underlying unidimensional structure. As a result, we excluded the positively worded items and tested a bi-factor model with 4 group factors again. The model had acceptable fit (CFI = .925, RMSEA = .088). In evaluation of local misfit, all residual correlations in the bi-factor model were below .2, thus supporting the fit of this model. The item loadings for the overall health impact factor were again higher (item loadings range .752–.952) than the item loadings in the four group factors which, respectively, were for family life from .200 to .524, for social life from .209 to .405, for occupational life from .004 to .421, for the generic factor from .056 to .435.

To test whether the acceptable fit of the bi-factor model was due solely to the removal of the positively worded items, we reran model 1a, excluding the positively worded items. While all residual correlations were below .2, the overall fit of the reduced unidimensional model was not acceptable (Table 2). We therefore regard model 3c as the best description of the data. In the process of model building, 11 items were excluded due to low item loadings, high residual correlations or poor fit.

Models including items on partner and children (Models 4a–4d)—In the subsample of participants who answered questions in the family domain regarding roles related to partner and children, the results on fit mimicked the previous results. The bi-factor model excluding positive items provided the best fit in this subsample as well (Table 1).

Discussion

We used an investigative approach with previously suggested modeling strategies to determine the factorial structure of a new role functioning item bank. We found that a bi-factor model with group factors for family, work, social, and generic function had the best fit. In this model, loadings on the general factor were considerably higher than the loadings on the group factors. These results suggest that the item bank is sufficiently unidimensional

for IRT analysis [12, 28–30]. Further, when comparing the loadings on the general impact factor in the bi-factor model and the loadings in the unidimensional model (Table 2), we find only negligible differences (mean = −.016, range −.017 to .049), providing additional support to the sufficient unidimensionality of the item bank. Thus, we conclude that the role functioning item bank is sufficiently unidimensional for IRT modeling. However, the remaining covariation between items that can be explained by the additional group factors for family, occupation and social life, suggest that content balancing of these subdomains may be useful in test construction.

We intentionally included some items with positive wording in an attempt to expand the measurement range of the bank, allowing it to assess even minor role functioning disruptions. Previous research on self-assessments including both positively and negatively worded items has suggested such measures can be interpreted as having one underlying latent factor and methods effects reflecting different item wording. The approach has been most extensively discussed in the psychometric literature on the Self-Esteem Scale [31], which in early publications has been presented as having two substantive dimensions based on EFA results [32, 33]. However, in subsequent analyses by CFA approaches, a unidimensional model with methods effects [34] has been proposed as the superior model. Subsequent studies have supported these solutions, even though various degrees of interpretability were assigned to the methods factors [19, 35, 36]. CFA modeling with methods effects has also been applied to scales measuring worry [26], physical anxiety [37], and self-concept [38] with similar conclusions.

In our study, modeling positively worded items improved item fit in the CFA framework, but the model showed very high loadings for the positive items on the methods factors, and large differences in the item slopes from the unidimensional model. As a result, we excluded these items from the item bank. It is worth pointing out that previous studies on methods effects were performed with relatively short (no more than 15 items), unidimensional measures. The large number of items in the bank, as well as the presence of several subdomains, could have impacted the final results in our study. In addition, the decision to include health attribution in all items may also be accountable for the problems with retaining positively worded items.

These specifics of testing our item bank—use of health attribution in all items, inclusion of several related domains of role functioning and large number of items—had broader implications than the exclusion of positively worded items from the final bank. Therefore, we will discuss each of these decisions and their implications below.

Early in the item development process we decided to include health attribution in all the items. Role functioning is a very broad concept that encompasses a variety of indicators and factors that can influence the level of functioning (e.g. opportunity, choice, interest, health). This conceptual breadth has important psychometric implications, including the appropriateness of use of unidimensional versus multidimensional model [14]. We believe it was important to include health attribution in the items in order to narrow down the breath of the measured construct and confine it within the realms of health-related quality of life. While this decision made the construct more narrowly defined, it also had some negative consequences. Health attributions made items longer and more cumbersome. In addition the formulation of positively worded items, describing unaltered or improved role functioning due to health through the use of health attribution presented some challenges in the readability and logic of formulated items. Results of the current study suggested that such items could not be retained in the bank.

Finally, the exploration of the factorial structure and sufficient unidimensionality of an item bank presumes the use of large number of items. It has been pointed out that in these situations, the use of standards of fit indices established with short measures is impractical since the large number of items typical for item banks and the skewness of the data (both conditions present in our data) lead to larger errors and worse fit index values [13]. While the sensitivity of fit indices to factors other than the dimensionality of the data has been demonstrated, to our knowledge there have been no specific alternative cut-off values proposed. An investigative approach and use of bi-factor models has been suggested instead as a more conceptually viable approach for the evaluations of unidimensionality of item banks [13, 14]. In our study, the best fit was demonstrated by bi-factor Model 3a, which had an acceptable fit even though fit indices were slightly below the universally recommended values.

In summary, we explored the factorial structure of an item bank assessing health impact on role functioning and found the bank to be sufficiently unidimensional for IRT applications. The limitations of this study include the skewness of the data, the presence of systematically missing data for some of the items, and the presence of ceiling effect for some participants. The next steps of the item bank development will be the calibration of the items on a common metric and the development of a computerized test of health impact on role functioning.

Acknowledgments

References

1. World Health Organization. Towards a common language for functioning, disability and health: ICF—the international classification of functioning, disability and health. 2002.

2. Badley EM. Enhancing the conceptual clarity of the activity and participation components of the international classification of functioning, disability, and health. Social Science and Medicine. 2008; 66:2335–2345. [PubMed: 18314239]

3. Merikangas KR, Ames M, Cui L, Stang PE, Ustun TB, Von Korff M, et al. The impact of comorbidity of mental and physical conditions on role disability in the US adult household population. Archives of General Psychiatry. 2007; 64:1180–1188. [PubMed: 17909130]

4. Druss BG, Hwang I, Petukhova M, Sampson NA, Wang PS, Kessler RC. Impairment in role functioning in mental and chronic medical disorders in the united states: Results from the national comorbidity survey replication. Molecular Psychiatry. 2009; 14:728–737. [PubMed: 18283278]

5. Anatchkova MD, Bjorner JB. Health and role functioning: The use of focus groups in the development of an item bank. Quality of Life Research. 2010; 19:111–123. [PubMed: 20047086]

6. Gandek B, Sinclair SJ, Jette AM, Ware JE Jr. Development and initial psychometric evaluation of the participation measure for post-acute care (PM-PAC). American Journal of Physical Medicine and Rehabilitation. 2007; 86:57–71. [PubMed: 17033591]

7. Haley SM, Gandek B, Siebens H, Black-Schaffer RM, Sinclair SJ, Tao W, et al. Computerized adaptive testing for follow-up after discharge from inpatient rehabilitation: II. Participation outcomes. Archives of Physical Medicine and Rehabilitation. 2008; 89:275–283. [PubMed: 18226651]

8. Mulcahey MJ, Haley SM, Duffy T, Pengsheng N, Betz RR. Measuring physical functioning in children with spinal impairments with computerized adaptive testing. Journal of Pediatric Orthopaedics. 2008; 28:330–335. [PubMed: 18362799]

9. Wilkie DJ, Judge MK, Berry DL, Dell J, Zong S, Gilespie R. Usability of a computerized PAINReportIt in the general public with pain and people with cancer pain. Journal of Pain and Symptom Management. 2003; 25:213–224. [PubMed: 12614956]

10. Bjorner JB, Kosinski M, Ware JE Jr. Calibration of an item pool for assessing the burden of headaches: An application of item response theory to the headache impact test (HIT). Quality of Life Research. 2003; 12:913–933. [PubMed: 14651412]

11. Bayliss MS, Dewey JE, Dunlap I, Batenhorst AS, Cady R, Diamond ML, et al. A study of the feasibility of internet administration of a computerized health survey: The headache impact test (HIT). Quality of Life Research. 2003; 12:953–961. [PubMed: 14651414]

12. Lai JS, Crane PK, Cella D. Factor analysis techniques for assessing sufficient unidimensionality of cancer related fatigue. Quality of Life Research. 2006; 15:1179–1190. [PubMed: 17001438]

13. Cook KF, Kallen MA, Amtmann D. Having a fit: Impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. Quality of Life Research. 2009; 18:447–460. [PubMed: 19294529]

14. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Quality of Life Research. 2007; 16(Suppl 1):19–31. [PubMed: 17479357]

15. O'Connor BP. SPSS and SAS programs for determining the number of components using parallel analysis and velicer's MAP test. Behavior Research Methods Instruments Computers. 2000; 32:396–402.

16. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, 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:S22–S31. [PubMed: 17443115]

17. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology. 2003; 88:879–903. [PubMed: 14516251]

18. Hankins M. The factor structure of the twelve item general health questionnaire (GHQ-12): The result of negative phrasing? Clinical Practice of Epidemiology in Mental Health. 2008; 4:10.

19. Horan PM, DiStefano C, Motl RW. Wording effects in self-esteem scales: Methodological artifact or response style? Structural Equation Modeling. 2003; 10:435–455.

20. Ware, JE., Jr; Dewey, J. How to score version two of the SF-36 health survey. Lincoln, RI: QualityMetric Incorporated; 2000.

21. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: Item banking and computerized adaptive assessment. Quality of Life Research. 2007; 16:95–108. [PubMed: 17530450]

22. Turner-Bowker DM, Anatchkova MD, Bjorner JB, Saris-Baglama RN, Chan KS, Huang I, Wu A. Preliminary development of a computerized adaptive test for health-related quality of life outcomes in HIV. The Quality of Life Research Journal. 2009

23. Marsh, HW.; Grayson, D. Latent variable models of multitrait—multimethod data. In: Hoyle, RH., editor. Structural equation modeling: Concepts, issues, and applications. Thousand Oaks, CA: Sage; 1995.

24. Muthen, B.; Muthen, L. Mplus user's guide. Los Angeles, CA: Muthen & Muthen; 1998.

25. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling. 1999; 6:1–55.

26. Brown TR. Confirmatory factor analysis of the penn state worry questionnaire: Multiple factors or method effects? Behaviour Research and Therapy. 2003; 41:1411–1426. [PubMed: 14583411]

27. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychological Methods. 1996; 1:130–149.

28. McDonald, RP. Test theory: A unified treatment. Mahwah, NJ: Lawrence Erlbaum Associates Inc; 1999.

29. Chen FF, West SG, Sousa KH. A comparison of bifactor and second-order models of quality of life. Multivariate Behavioral Research. 2006; 41:189–225.

30. Chen WH, Revicki DA, Lai JS, Cook KF, Amtmann D. Linking pain items from two studies onto a common scale using item response theory. Journal of Pain and Symptom Management. 2009; 38:615–628. [PubMed: 19577422]

31. Rosenberg, M. Society and the adolescent child. Princeton, NJ: Princeton University Press; 1965.

32. Bachman JG, O'Malley PM. Self-concepts, self-esteem, and educational experiences: The frog pond revisited (again). Journal of Personality and Social Psychology. 1986; 50:35–46.

33. Carmines, EG.; Zeller, RA. Reliability and validity assessment. Beverly Hills, CA: Sage; 1979.

34. Marsh HW. Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? Journal of Personality and Social Psychology. 1996; 70:810–819. [PubMed: 8636900]

35. Tomas JM, Oliver A. Rosenberg's self-esteem scale: Two factors or method effects. Structural Equation Modeling. 1999; 6:84–98.

36. Wang J, Siegal HA, Falck RS, Carlson RG. Factorial structure of Rosenberg's self-esteem scale among crack-cocaine drug users. Structural Equation Modeling. 2001; 8:275–286.

37. Motl RW, Conroy DE. Validity and factorial invariance of the social physique anxiety scale. Medicine and Science in Sports and Exercise. 2000; 32:1007–1017. [PubMed: 10795794]

38. Chen, Y.; Rendina-Gobioff, G.; Dedrick, RF. Detecting effects of positively and negatively worded items on a self-concept scale for third and sixth grade elementary students. Paper presented at the Annual Meeting of the Florida Educational Research Association; Tampa, FL. Nov 14–16, 2007; 2007. 52nd
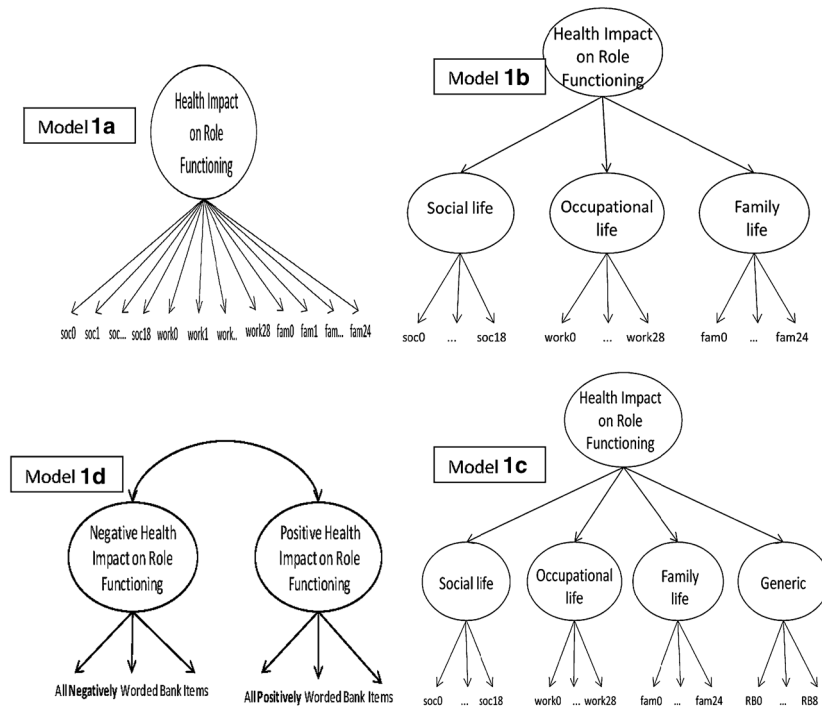
Fig. 1.
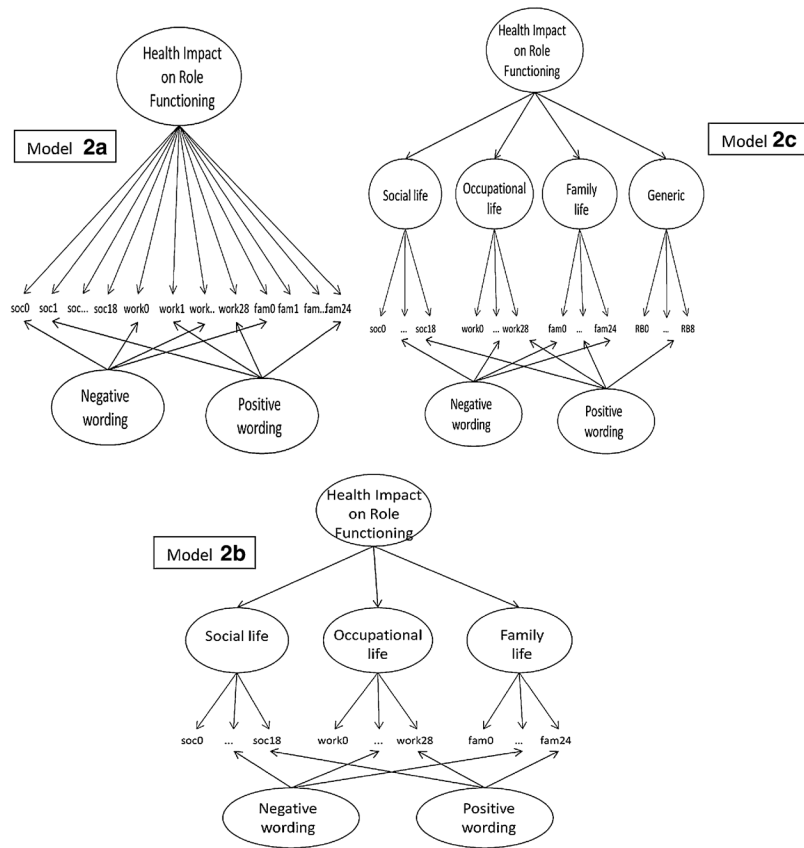Health impact concepts models (Models 1a–1d)

Fig. 2.
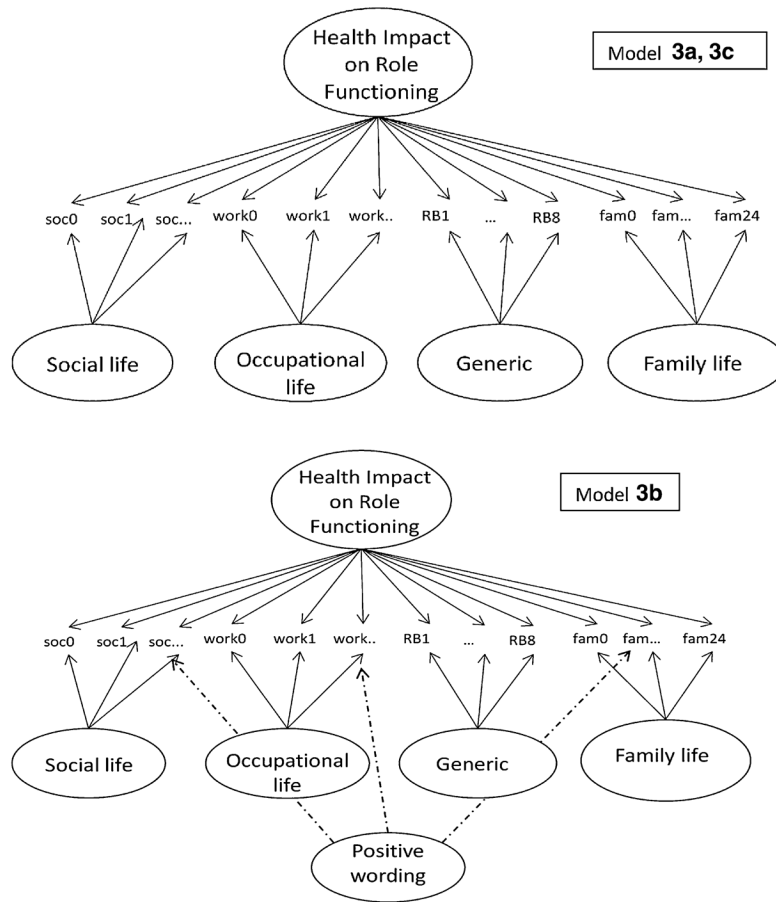Health impact concepts and methods components models (Models 2a–2c)

Fig. 3.
Bi-factor models (Models 3a–3c). *Note*: model 3c has the same factor structure as model 3a, but excludes positively worded items

Table 1

Fit indexes of CFA models of health impact on role functioning item bank

Model	Description	Fit indexes			
		Chi-square	df	CFI	RMSEA
Models excluding items on partner and children, n = 1193					
Health impact concepts models					
Model 1a	1 concept (Health impact on RF)	3247.75	106	0.840	0.158
Model 1b	3 concepts (Hierarchical model of health impact on social, family, occupational life)	2844.15	126	0.861	0.134
Model 1c	4 concepts (Hierarchical model of health impact on social, family, occupational life, generic impact)	2651.64	141	0.872	0.122
Model 1d	2 concepts (Negative and Positive Health Impact)	2399.08	110	0.883	0.132
Health impact concepts and methods component models					
Model 2a	1 concept, 2 methods	2371.98	110	0.884	0.131
Model 2b	3 concepts, 2 methods (Hierarchical)	1935.91	133	0.908	0.107
Model 2c	4 concepts, 2 methods (Hierarchical)	1609.21	152	0.925	0.090
Bi-factor models					
Model 3a	1 general, 4 group factors	2203.22	149	0.895	0.108
Model 3b	1 general, 4 group, 1 positive methods factors	1459.03	154	0.933	0.084
Model 3c	**1 general, 4 group factors, positive items excluded**	**1447.51**	**140**	**0.925**	**0.088**
Models including items on partner and children N = 414					
Model 4a	1 concept, 2 methods	629.4	88	0.929	0.122
Model 4b	3 concepts, 2 methods (Hierarchical)	489.83	96	0.949	0.100
Model 4c	4 concepts, 2 methods (Hierarchical)	427.53	103	0.958	0.087
Model 4d	**Bi-factor model 1 general, 4 group factors, positive items excluded**	**423.22**	**108**	**0.959**	**0.084**

Final models retained are presented in bold font. Decision was based on examination of patterns of standardized item loadings, CFI >.900 and RMSEA <.09

Table 2

Parameter estimates for unidimensional and bi-factor models

			Single factor (with positive items, Model 1a)	Single factor (without positive items)	Bi-factor model (Model 3c)				
CFI			**0.840**	**0.873**	**0.925**				
RMSEA			**0.158**	**0.140**	**0.088**				
Item	**Abbreviated item text**	**Response**			**Factor loadings**				
			Unidimensional		Overall impact	Family	Work	Social	Generic
RB1C	…because of health …ability to perform activities such as housework or chores reduced?	A	0.913	0.914	0.912				0.18
RB2C	…because of health …ability to engage in non-work activities reduced?	A	0.943	0.944	0.952				0.099
RB3C	…because of health …ability to work (at home, school, or the workplace) reduced?	A	0.927	0.927	0.929				0.152
RB4cc	…because of health …had difficulty in performing work or other daily activities?	B	0.946	0.947	0.949				0.147
RB5cc	…health keep you from getting as much done at work or at home?	B	0.946	0.947	0.951				0.134
RB6c	…health limit your ability to concentrate on work or daily activities?	B	0.906	0.906	0.919				0.056
RB7	…because of health need help in handling routine tasks such as every day household chores, doing necessary business, shopping, or caring for others, because of your health?	B	0.858	0.859	0.862				0.158
RB8c	… because of health have to stop work or other activities?	B	0.891	0.893	0.896				0.17
RP1	As result of physical health …cut down on the amount of time spent on work or other activities	B	0.911	0.911	0.888				0.336
RP2	As result of physical health …accomplished less than you would like	B	0.9	0.901	0.881				0.33
RP3	As result of physical health …limited in the kind of work or other activities	B	0.902	0.903	0.857				0.435
RP4	As result of physical health …had difficulty performing the work or other activities	B	0.924	0.924	0.883				0.413
Fam0	…health have a negative impact on your family life?	C	0.818	0.820	0.817	0.219			
Fam1	…health make you worry about providing for your family?		0.752	0.757	0.705	0.524			
Fam2	…health cause family conflicts?	C	0.751	0.753	0.721	0.414			

Model description		Single factor (with positive items, Model 1a)	Single factor (without positive items)	Bi-factor model (Model 3c)		
CFI		0.840	0.873	0.925		
RMSEA		0.158	0.140	0.088		

Item	Abbreviated item text	Response	Factor loadings					
			Unidimensional	Overall impact	Family	Work	Social	Generic
Fam3	...did your health make you worry about the future of your family?	C	0.776	0.780	0.508			
Fam4	...health make you irritable toward family members?	D	0.727	0.731	0.205			
Fam5	...avoid talking to family members because of your health?	D	0.774	0.776	0.312			
Fam6	...health make it hard for you to do chores?	D	0.87	0.873	0.206			
Fam7	...health make it hard to meet family responsibilities?	C	0.87	0.872	0.259			
Fam8	...health limit your ability to give emotional support to your family?	C	0.797	0.799	0.257			
Fam9	...health make driving difficult ?	C	0.789	0.791	0.225			
Fam10	...health make it difficult to provide for your family ?	C	0.855	0.857	0.305			
Fam11	...health make it impossible to do housework?	D	0.858	0.860	0.229			
Fam12	...could not support my family because of my health.		0.864	0.865	0.43			
Fam13	...health make it difficult for you to take care of family members?	D	0.886	0.889	0.37			
Fam14	...could not take care of my family because of my health.		0.856	0.858	0.454			
Fam15	...family members have to do some of your household tasks for you?	D	0.853	0.855	0.274			
Fam16	...health make you feel entirely dependent on family members?	C	0.84	0.841	0.283			
Fam17	...feel like a burden to your family because of your health?	D	0.855	0.856	0.275			
Fam20[b]	...worry that...partner relationship may suffer because of health?	D	0.779	0.780	0.28			
Fam21C[b]	...avoid talking to partner because of health?	D	0.802	0.804	0.382			
Fam22C[b]	...make unreasonable demands on partner because of health?	C	0.858	0.859	0.323			
Fam23C[b]	...need partner to take care of you because of your health?	D	0.832	0.834	0.33			
Fam26[b]	...find it hard to comfort your children because of your health?	D	0.732	0.737	0.278			
Fam27C[b]	...health make it hard to take care of your children?	C	0.897	0.899	0.345			

			Single factor (with positive items, Model 1a)	Single factor (without positive items)	Bi-factor model (Model 3c)				
CFI			**0.840**	**0.873**	**0.925**				
RMSEA			**0.158**	**0.140**	**0.088**				
Item	**Abbreviated item text**	**Response**			**Factor loadings**				
			Unidimensional		Overall impact	Family	Work	Social	Generic
Fam28[b]	…feel you could not support your children because of health?	D	0.904	0.774	0.72	0.418			
Soc0	…health impact your social life in the past 4 weeks?	C	0.914	0.916	0.866			0.401	
Soc1	…health limit your enjoyment in social activities?	C	0.923	0.924	0.874			0.405	
Soc2	…health prevent you from being with friends?	D	0.908	0.909	0.885			0.284	
Soc3	…lose your enjoyment in leisure activities because of your health?	C	0.891	0.893	0.872			0.28	
Soc4	…lose your enjoyment from social events because of health?	C	0.912	0.913	0.883			0.322	
Soc5	…health lessen your desire to get together with friends?	C	0.912	0.912	0.875			0.351	
Soc6	…limit your community involvement because of your health?	D	0.89	0.891	0.86			0.328	
Soc7	…health limit you in dining out?	C	0.823	0.825	0.805			0.294	
Soc8	…health make it hard to find time for your friends ?	C	0.919	0.919	0.883			0.343	
Soc9	…health make you feel you had no energy to keep in touch with friends?	C	0.876	0.878	0.853			0.303	
Soc10	…have little energy for your hobbies because of your health?	D	0.802	0.804	0.796			0.209	
Soc11	…hard to keep up with your social responsibilities?	C	0.912	0.914	0.893			0.274	
Soc12	…health limit your participation in activities you enjoy?	C	0.932	0.932	0.917			0.239	
Soc14	…have to change your social activities because of your health?	C	0.902	0.902	0.882			0.277	
Soc15[a]	…health allow you to enjoy your social life?	C	0.68	Positive item excluded in model					
Soc16[a]	…health allow you to stay actively involved in the community?	C	0.743	Positive item excluded in model					
Soc17[a]	…health allow you to meet all of your social responsibilities?	C	0.728	Positive item excluded in model					
Work0	…health have any impact on your work?	C	0.905	0.906	0.905		0.149		
Work1	…did you worry that your health may interfere with your work?	C	0.898	0.899	0.904		0.112		
Work2	…did you feel nervous at work because of your health?	D	0.866	0.867	0.865		0.173		
Work3	…did your health make you irritable at work?	D	0.832	0.833	0.812		0.308		

Model description			Single factor (with positive items, Model 1a)	Single factor (without positive items)	Bi-factor model (Model 3c)				
CFI			**0.840**	**0.873**	**0.925**				
RMSEA			**0.158**	**0.140**	**0.088**				
					Factor loadings				
Item	Abbreviated item text	Response	Unidimensional		Overall impact	Family	Work	Social	Generic
Work4	…because of health did you experience more stress than usual at work?	C	0.869	0.870	0.858		0.236		
Work5	…health make work demands overwhelming for you?	D	0.895	0.895	0.887		0.215		
Work6	…health cause financial worries?	D	0.837	0.839	0.851		0.034		
Work7	…health make it hard to focus on your work?	D	0.917	0.918	0.903		0.26		
Work8	…health make it hard to talk to your colleagues?	C	0.839	0.841	0.812		0.345		
Work9	…because of health problems communicating with your colleagues and supervisors?	D	0.86	0.861	0.831		0.356		
Work10	…health make it hard to start work on time?	C	0.816	0.817	0.804		0.247		
Work11	…health make you fall behind on your work?	C	0.883	0.885	0.854		0.346		
Work12	…produce less at work because of your health?	C	0.898	0.899	0.855		0.394		
Work13	…health lower your work quality?	C	0.875	0.876	0.832		0.401		
Work14	…health make it physically hard to get through the work day?	C	0.908	0.909	0.911		0.144		
Work15	…health make you less productive at work than others in a similar position?	C	0.873	0.874	0.826		0.421		
Work16	…health make it hard to concentrate on your work?	C	0.893	0.894	0.875		0.278		
Work17	…health force you to work extra hours to keep up with the workload?	D	0.764	0.766	0.747		0.286		
Work18	…make more mistakes at work because of your health?	D	0.834	0.836	0.804		0.357		
Work19	…work performance suffer because of your health?	C	0.895	0.897	0.85		0.41		
Work20	…colleagues and supervisor express concerns that your health affected the quality of your work?	D	0.751	0.753	0.719		0.372		
Work21	…health cause you difficulties getting to and from work?	C	0.814	0.816	0.818		0.138		
Work23	…work slower than usual because of your health?	C	0.882	0.882	0.873		0.222		
Work24	…health cause you financial difficulties?	C	0.809	0.811	0.826		0.004		
Work27	…worry about losing your job because of your health?	D	0.798	0.799	0.793		0.196		
Work28[a]	…health allow you to perform at your best?	C	0.74	Positive item excluded in model					

Model description	Single factor (with positive items, Model 1a)	Single factor (without positive items)	Bi-factor model (Model 3c)					
CFI	**0.840**	**0.873**	**0.925**					
RMSEA	**0.158**	**0.140**	**0.088**					

Item	Abbreviated item text	Response	Factor loadings					
			Unidimensional	Overall impact	Family	Work	Social	Generic
Work29[a]	…health allow you to feel competitive at your job?	C	0.703	Positive item excluded in model				
Items excluded from reported models								
Fam18	…health allow you to provide for your family?	D	Item was excluded in initial exploratory CFA's due to low loading and high residula correlations with other items					
Fam19	…health allow you to provide emotional support to your partner?	C	Item was excluded in initial exploratory CFA's due to low loading and high residula correlations with other items					
Fam24	…health allow you to take care of your children?	D	Positive item excluded from ananlyses in SS2					
Fam25	…health allow you to play with your children?	C	Positive item excluded from ananlyses in SS3					
Soc13	…health make it hard to participate in your favorite leisure activities?	D	Item was excluded in initial exploratory CFA's due to low loading and high residula correlations with other items					
Work22	…have the physical energy needed for your work?	D	Item was administered, but it had no health attribution and was assessed to be more relevant to physical functioning.					

Items with RB labels are from previously developed short role functioning item bank. A "C" following a item label indicates that the item has collapsed response options

Response formats: A Not reduced at all, Reduced a little, Somewhat reduced, Reduced very much, Unable to perform; *B* None of the time, A little of the time, Some of the time, A little of the time, Some of the time, Most of the time, All the time; *C* Not at all, A little, Somewhat, Quite a lot, Extremely; *D* Always, Often, Sometimes, Rarely, Never

[a] Positively worded items excluded from the model

[b] Estimates for these items are based on the models with smaller sample size (Model 4d)