Incorporating different response formats of competence tests in an IRT model
Kerstin Haberkorn, Steffi Pohl & Claus H. Carstensen
Abstract | PDF of the full article

In memoriam Benjamin Wright (1926 – 2015)
David Andrich PDF of the full article

Measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) Applied Cognition General Concerns, short forms in ethnically diverse groups
Robert Fieo, Katja Ocepek-Welikson, Marjorie Kleinman, Joseph P. Eimicke, Paul K. Crane, David Cella & Jeanne A. Teresi
Abstract | PDF of the full article

Measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) Pain Interference short form items: Application to ethnically diverse cancer and palliative care populations
Jeanne A. Teresi, Katja Ocepek-Welikson, Karon F. Cook, Marjorie Kleinman, Mildred Ramirez, M. Carrington Reid & Albert Siu
Abstract | PDF of the full article

Measurement properties of PROMIS Sleep Disturbance short forms in a large, ethnically diverse cancer cohort
Roxanne E. Jensen, Bellinda L. King-Kallimanis, Eithne Sexton, Bryce B. Reeve, Carol M. Moinpour, Arnold L. Potosky, Tania Lobo & Jeanne A. Teresi
Abstract | PDF of the full article

Differential item functioning in Patient Reported Outcomes Measurement Information System® (PROMIS®) Physical Functioning short forms: Analyses across ethnically diverse groups
Richard N. Jones, Doug Tommet, Mildred Ramirez, Roxanne Jensen & Jeanne A. Teresi
Abstract | PDF of the full article

Measuring social function in diverse cancer populations: Evaluation of measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) Ability to Participate in Social Roles and Activities Short Form
Elizabeth A. Hahn, Michael A. Kallen, Roxanne E. Jensen, Arnold L. Potosky, Carol M. Moinpour, Mildred Ramirez, David Cella & Jeanne A. Teresi
Abstract | PDF of the full article
Incorporating different response formats of competence tests in an IRT model
Kerstin Haberkorn, Steffi Pohl & Claus H. Carstensen

Abstract

Competence tests within large-scale assessments usually contain various task formats to measure the participants' knowledge. Two response formats that are frequently used are simple multiple choice (MC) items and complex multiple choice (CMC) items. Whereas simple MC items comprise a number of response options with one being correct, CMC items consist of several dichotomous true-false subtasks. When incorporating these response formats in a scaling model, they are mostly assumed to be unidimensional. In empirical studies different empirical and theoretical schemes of weighting CMC items in relation to MC items have been applied to construct the overall competence score. However, the dimensionality of the two response formats and the different weighting schemes have only rarely been evaluated. The present study, thus, addressed two questions of particular importance when implementing MC and CMC items in a scaling model: Do the different response formats form a unidimensional construct and, if so, which of the weighting schemes considered for MC and CMC items appropriately models the empirical competence data? Using data of the National Educational Panel Study, we analyzed scientific literacy tests embedding MC and CMC items. We cross-validated the findings on another competence domain and on another large-scale assessment. The analyses revealed that the different response formats form a unidimensional measure across contents and studies. Additionally, the a priori weighting scheme of one point for MC items and half points for each subtask of CMC items best modeled the response formats impact on the competence score and resembled the empirical competence data well.

Key words: item response theory, complex multiple choice, item format weighting, scoring, dimensionality

Dr. Kerstin Haberkorn
Breslaustr. 12
96052 Bamberg, Germany
kerstin.haberkorn@uni-bamberg.de

Measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) Applied Cognition – General Concerns, short forms in ethnically diverse groups
Robert Fieo, Katja Ocepek-Welikson, Marjorie Kleinman, Joseph P. Eimicke, Paul K. Crane, David Cella & Jeanne A. Teresi

Abstract

Aims: The goals of these analyses were to examine the psychometric properties and measurement equivalence of a self-reported cognition measure, the Patient Reported Outcome Measurement Information System® (PROMIS®) Applied Cognition General Concerns short form. These items are also found in the PROMIS Cognitive Function (version 2) item bank. This scale consists of eight items related to subjective cognitive concerns. Differential item functioning (DIF) analyses of gender, education, race, age, and (Spanish) language were performed using an ethnically diverse sample (n = 5,477) of individuals with cancer. This is the first analysis examining DIF in this item set across ethnic and racial groups.

Methods: DIF hypotheses were derived by asking content experts to indicate whether they posited DIF for each item and to specify the direction. The principal DIF analytic model was item response theory (IRT) using the graded response model for polytomous data, with accompanying Wald tests and measures of magnitude. Sensitivity analyses were conducted using ordinal logistic regression (OLR) with a latent conditioning variable. IRT-based reliability, precision and
information indices were estimated.

Results: DIF was identified consistently only for the item, brain not working as well as usual. After correction for multiple comparisons, this item showed significant DIF for both the primary and sensitivity analyses. Black respondents and Hispanics in comparison to White non-Hispanic respondents evidenced a lower conditional probability of endorsing the item, brain not working as well as usual. The same pattern was observed for the education grouping variable: as compared to those with a graduate degree, conditioning on overall level of subjective cognitive concerns, those with less than high school education also had a lower probability of endorsing this item. DIF was observed for age for two items after correction for multiple comparisons for both the IRT and OLR-based models: I have had to work really hard to pay attention or I would make a mistake and I have had trouble shifting back and forth between different activities that require thinking. For both items, conditional on cognitive complaints, older respondents had a higher likelihood than younger respondents of endorsing the item in the cognitive complaints direction. The magnitude and impact of DIF was minimal.

The scale showed high precision along much of the subjective cognitive concerns continuum; the overall IRT-based reliability estimate for the total sample was 0.88 and the estimates for subgroups ranged from 0.87 to 0.92.

Conclusion: Little DIF of high magnitude or impact was observed in the PROMIS Applied Cognition – General Concerns short form item set. One item, “It has seemed like my brain was not working as well as usual” might be singled out for further study. However, in general the short form item set was highly reliable, informative, and invariant across differing race/ethnic, educational, age, gender, and language groups.

Key words: PROMIS®, cognitive concerns, item response theory, differential item functioning, race, ethnicity

Robert Fieo, Assistant Professor
University of Florida
College of Medicine
Department of Geriatric Research
2004 Mowry Road
Gainesville, FL 32611, USA
fieo@ufl.edu

Measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) Pain Interference short form items: Application to ethnically diverse cancer and palliative care populations

Jeanne A. Teresi, Katja Ocepek-Welikson, Karon F. Cook, Marjorie Kleinman, Mildred Ramirez, M. Carrington Reid & Albert Siu

Abstract
Reducing the response burden of standardized pain measures is desirable, particularly for individuals who are frail or live with chronic illness, e.g., those suffering from cancer and those in palliative care. The Patient Reported Outcome Measurement Information System® (PROMIS®) project addressed this issue with the provision of computerized adaptive tests (CAT) and short form measures that can be used clinically and in research. Although there has been substantial evaluation of PROMIS item banks, little is known about the performance of PROMIS short forms, particularly in ethnically diverse groups. Reviewed in this article are findings related to the differential item functioning (DIF) and reliability of the PROMIS pain interference short forms across diverse socio-demographic groups.

Methods: DIF hypotheses were generated for the PROMIS short form pain interference items. Initial analyses tested item response theory (IRT) model assumptions of unidimensionality and local independence. Dimensionality was evaluated using factor analytic methods; local dependence (LD) was tested using IRT-based LD indices. Wald tests were used to examine group differences in IRT parameters, and to test DIF hypotheses. A second DIF-detection method used in sensitivity analyses was based on ordinal logistic regression with a latent IRT-derived conditioning variable. Magnitude and impact of DIF were investigated, and reliability and item and scale information statistics were estimated.

Results: The reliability of the short form item set was excellent. However, there were a few items with high local dependency, which affected the estimation of the final discrimination parameters. As a result, the item, “How much did pain interfere with enjoyment of social activities?” was excluded in the DIF analyses for all subgroup comparisons. No items were hypothesized to show DIF for race and ethnicity; however, five items showed DIF
after adjustment for multiple comparisons in both primary and sensitivity analyses: ability to concentrate, enjoyment of recreational activities, tasks away from home, participation in social activities, and socializing with others. The magnitude of DIF was small and the impact negligible. Three items were consistently identified with DIF for education: enjoyment of life, ability to concentrate, and enjoyment of recreational activities. No item showed DIF above the magnitude threshold and the impact of DIF on the overall measure was minimal. No item showed gender DIF after correction for multiple comparisons in the primary analyses. Four items showed consistent age DIF: enjoyment of life, ability to concentrate, day to day activities, and enjoyment of recreational activities, none with primary magnitude values above threshold. Conditional on the pain state, Spanish speakers were hypothesized to report less pain interference on one item, enjoyment of life. The DIF findings confirmed the hypothesis; however, the magnitude was small.

Using an arbitrary cutoff point of theta (θ) ≥ 1.0 to classify respondents with acute pain interference, the highest number of changes were for the education groups analyses. There were 231 respondents (4 % of the total sample) who changed from the designation of no acute pain interference to acute interference after the DIF adjustment. There was no change in the designations for race/ethnic subgroups, and a small number of changes for respondents aged 65 to 84.

Conclusions: Although significant DIF was observed after correction for multiple comparisons, all DIF was of low magnitude and impact. However, some individual-level impact was observed for low education groups. Reliability estimates were high. Thus, the PROMIS short form pain items examined in this ethnically diverse sample performed relatively well; although one item was problematic and removed from the analyses. It is concluded that the majority of the PROMIS pain interference short form items can be recommended for use among ethnically diverse groups, including those in palliative care and with cancer and chronic illness.

Key words: Differential item functioning, PROMIS®, pain, measurement equivalence, palliative care, ethnicity, cancer

Jeanne A. Teresi, Ed.D, Ph.D.
Columbia University Stroud Center
at New York State Psychiatric Institute
1051 Riverside Drive, Box 42, Room 2714, New York
New York, 10032-3702, USA
Teresimeas@aol.com

Measurement properties of PROMIS Sleep Disturbance short forms in a large, ethnically diverse cancer cohort
Roxanne E. Jensen, Bellinda L. King-Kallimanis, Eithne Sexton, Bryce B. Reeve, Carol M. Moinpour, Arnold L. Potosky, Tania Lobo & Jeanne A. Teresi

Abstract
AIMS: To evaluate model fit, differential item function (DIF), and construct validity of select short forms from the PROMIS® Sleep Disturbance item bank.
METHODS: We recruited cancer survivors who were between 6 - 13 months post diagnosis (n = 4,956), as part of the Measuring Your Health (MY-Health) study. We measured sleep disturbance using 10 items commonly found in PROMIS Sleep Disturbance short forms (Sleep 4a, Sleep 6a, Sleep 8b), and which are frequently administered in computerized adaptive testing. We evaluated domain reliability using Cronbach's coefficient alpha and factorial validity by fitting a PROMIS Sleep Disturbance unidimensional measurement model using confirmatory factor analysis (CFA). At the item-level, we examined DIF with respect to race/ethnicity (non-Hispanic White [NHW], non-Hispanic Black [NHB], Hispanic, and Asian/Pacific Islander), age, and sex. We used a multi-group CFA and multiple indicators, multiple methods (MIMIC) analyses. We then assessed construct validity (convergent, discriminate, and known groups) for sleep short forms, and a new best fit 6-item sleep disturbance short form.
RESULTS: We identified a satisfactory unidimensional sleep disturbance 6-item measure (χ²(6)37.6, p < 0.001, RMSEA = 0.031). To achieve this, we removed four items from the model with item content overlap and added residual covariances between positively worded items in order to address a method effect. We identified one instance of DIF: NHW participants were less likely to agree with the statement I had difficulty falling asleep compared to NHBs, Hispanics, or Asians/Pacific Islanders, who all reported the same level of sleep disturbance. After controlling for DIF, we extended this into a MIMIC model, identifying no additional DIF by age or sex. Across all race/ethnicity groups, the adjusted overall means suggest that older adults reported significantly lower sleep disturbance, and NHW, NHB, and Hispanic women reported significantly higher sleep...
disturbance than male survivors of the same race/ethnicity.

CONCLUSIONS: We could not fit a unidimensional measurement model for either the full 10-items, or for any combination of sleep disturbance items used in PROMIS Sleep Disturbance short forms. However, after we removed the overlapping item content and adjusted for methods effects, a 6-item measurement model for sleep disturbance fit the data well, with very little evidence of substantial DIF. This suggests this new measure (Sleep 6b) can be used in different groups across the adult lifespan, and in males and females in a heterogeneous cancer population. Our findings suggest further validation work is necessary to understand the impact of reverse-scored items, response set effects, and content overlap in this item bank.

Key words: sleep disturbance, PROMIS, differential item functioning, measurement invariance, methods effects

Roxanne E. Jensen, Ph.D.
Cancer Prevention and Control Program
Lombardi Comprehensive Cancer Center
Georgetown University
3300 Whitehaven Street NW, Suite 4100
Washington, DC 20007, USA
rj222@georgetown.edu

Differential item functioning in Patient Reported Outcomes Measurement Information System® (PROMIS®) Physical Functioning short forms: Analyses across ethnically diverse groups

Richard N. Jones, Doug Tommet, Mildred Ramirez, Roxanne Jensen & Jeanne A. Teresi

Abstract

We analyzed physical functioning short form items derived from the PROMIS® item bank (PF16) using data from more than 5,000 recently diagnosed, ethnically diverse cancer patients. Our goal was to determine if the short form items demonstrated evidence of differential item functioning (DIF) according to sociodemographic characteristics in this clinical sample. We evaluated responses for evidence of unidimensionality, local independence (given a single common factor), differential item functioning, and DIF impact. DIF was evaluated attributable to sex, age (middle aged vs. younger and older), race/ethnicity (White vs. Black or African-American, Asian/Pacific Islander, Hispanic) and level of education. We used a multiple group confirmatory factor analysis with covariates approach, a multiple indicators multiple causes (MIMIC) model. We confirmed essential unidimensionality but some evidence for multidimensionality is present, particularly for basic activities of daily living items, and many instances of local dependence. The presence of local dependence calls for further review of the meaning and measurement of the physical functioning domain among cancer patients.

Nearly every item demonstrated statistically significant DIF. In all group comparisons the impact of DIF was negligible. However, the Hispanic subgroup comparison revealed an impact estimate just below an arbitrary threshold for small impact. Within the limitations of local dependency violations, we conclude that items from a static short form derived from the PROMIS physical functioning item bank displayed trivial and ignorable DIF attributable to sex, race, ethnicity, age, and education among cancer patients.

Key words: PROMIS, physical function, differential item functioning, cancer, measurement equivalence, ethnicity

Richard N. Jones, Sc.D.
Department of Psychiatry and Human Behavior
Department of Neurology
Warren Alpert Medical School
Brown University, Butler Hospital
345 Blackstone Boulevard, Box G-BH, Providence
Rhode Island 02906, USA
richard_jones@brown.edu
Measuring social function in diverse cancer populations: Evaluation of measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) Ability to Participate in Social Roles and Activities short form

Elizabeth A. Hahn, Michael A. Kallen, Roxanne E. Jensen, Arnold L. Potosky, Carol M. Moinpour, Mildred Ramirez, David Cella & Jeanne A. Teresi

Abstract

Conceptual and psychometric measurement equivalence of self-report questionnaires are basic requirements for valid cross-cultural and demographic subgroup comparisons. The purpose of this study was to evaluate the psychometric measurement equivalence of a 10-item PROMIS® Social Function short form in a diverse population-based sample of cancer patients obtained through the Measuring Your Health (MY-Health) study (n = 5,301). Participants were cancer survivors within six to 13 months of a diagnosis of one of seven cancer types, and spoke English, Spanish, or Mandarin Chinese. They completed a survey on sociodemographic and clinical characteristics, and health status. Psychometric measurement equivalence was evaluated with an item response theory approach to differential item functioning (DIF) detection and impact. Although an expert panel proposed that many of the 10 items might exhibit measurement bias, or DIF, based on gender, age, race/ethnicity, and/or education, no DIF was detected using the study’s standard DIF criterion, and only one item in one sample comparison was flagged for DIF using a sensitivity DIF criterion. This item’s flagged DIF had only a trivial impact on estimation of scores. Social function measures are especially important in cancer because the disease and its treatment can affect the quality of marital relationships, parental responsibilities, work abilities, and social activities. Having culturally relevant, linguistically equivalent and psychometrically sound patient-reported measures in multiple languages helps to overcome some common barriers to including underrepresented groups in research and to conducting cross-cultural research.

Key words: patient-reported outcomes, social function, psychometrics, differential item functioning, cancer

Elizabeth A. Hahn, PhD
Department of Medical Social Sciences
Northwestern University Feinberg School of Medicine
633 N. St. Clair St., Suite 1900
Chicago, IL 60611, USA
e-hahn@northwestern.edu

Epilogue to the two-part series: Measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) short forms

Jeanne A. Teresi & Bryce B. Reeve

Abstract

The articles in this two-part series of Psychological Test and Assessment Modeling describe the psychometric performance and measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) short form measures in ethnically, socio-demographically diverse groups of cancer patients. Measures in eight health-related quality of life domains were evaluated: fatigue, depression, anxiety, cognition, pain, sleep, and physical and social function. State-of-the-art latent variable methods, most based on item response theory, and described in two methods overview articles in this series were used to examine differential item functioning (DIF).

Findings were generally supportive of the performance of the PROMIS measures. Although use of powerful methods and large samples resulted in the identification of many items with DIF, practically none were identified with high magnitude. The aggregate level impact of DIF was small, and minimal individual impact was detected. Some methodological challenges were encountered in-volving positively and negatively worded items, but most were resolved through modest item removal. Sensitivity analyses showed minimal impact of model assumption violation on the results presented.

A cautionary note is the observance of a few instances of individual-level impact of DIF in the analyses of depression, anxiety, and pain, and one instance of aggregate level impact just below threshold in the analyses of physical function. Although this sample of over 5,000 individuals was diverse, ethnically, a limitation was the lack of ability to examine language groups other than Spanish and English and specific ethnic subgroups within Hispanic, Asian/Pacific Islander, and Black subsamples.
Extensive qualitative and quantitative analyses were performed in the development of PROMIS item banks. These sets of analyses, performed by several teams of psychometricians, statisticians, and qualitative experts, were the first to examine measurement equivalence of PROMIS short forms among ethnically diverse groups, and were also the first examination of PROMIS short forms among adults with cancer. Results presented in these articles provide strong evidence supporting the measurement equivalence of PROMIS short forms.

Key words: Patient Reported Outcomes Measurement Information System®, PROMIS®, short forms, measurement equivalence, differential item functioning, item response theory, reliability, validity, ethnic diversity, cancer

Jeanne A. Teresi, Ed.D., Ph.D.
Columbia University Stroud Center
at New York State Psychiatric Institute
1051 Riverside Drive, Box 42, Room 2714
New York 10032-3702, USA
Teresimeas@aol.com