



Multidimensionality of the PROMIS self-efficacy measure for managing chronic conditions

Mi Jung Lee^{1,2} · Sergio Romero^{1,2} · Craig A. Velozo³ · Ann L. Gruber-Baldini⁴ · Lisa M. Shulman⁵

Accepted: 14 January 2019 / Published online: 26 February 2019
© Springer Nature Switzerland AG 2019

Abstract

Purpose This study investigated the PROMIS Self-Efficacy Measure for Managing Chronic Conditions (PROMIS-SE) domain distributions and examined the factor structure of the PROMIS-SE.

Methods A total of 1087 individuals with chronic conditions participated in this study. PROMIS-SE's item banks and two short forms (eight-item and four-item) measuring five behavioral domains (daily activities(DA), Emotions(EM), medications and treatments(MT), social interactions(SS), and Symptoms(SX)) were examined. PROMIS-SE's T-score ranges and distributions were examined to identify domain metric distributions and confirmatory factor analysis (CFA) was conducted to test a multidimensional model fit to the PROMIS-SE.

Results PROMIS-SE domains showed different T-score ranges and distributions for item banks and two short forms across all five domains. While PROMIS-SE EM demonstrated the highest T-scores (least negatively skewed), MT had the lowest T-scores (most negatively skewed) for all three forms. In general, respondents were more likely to achieve highest self-efficacy ratings (very confident) for domains DA, MT, and SS as compared to domains EM and SX. CFA confirmed that a multidimensional model adequately fit all three PROMIS-SE forms.

Conclusion Our results indicate that self-efficacy T-distributions are not consistent across domains (i.e., managing medications and treatments domain was more negatively skewed difficult than other domains), which is a requirement for making inter-domain comparisons. A multidimensional model could be used to enhance the PROMIS-SE's estimate accuracy and clinical utility.

Keywords Patient-reported outcome measure · Self-efficacy · PROMIS · Multidimensionality

Bandura defines self-efficacy as individuals' confidence in performing a required behavior or task to achieve a desired outcome [1]. To be specific, individuals' self-efficacy is not related to the skills they have, but their confidence in what

they can perform [2]. For example, a patient may have the skills needed to manage medications but may not be confident to do so.

Self-efficacy has consistently predicted the initiation of behavioral changes and maintenance of acquired health behaviors [3]. Several studies have demonstrated that self-efficacy is associated with individuals' health outcomes

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11136-019-02116-w>) contains supplementary material, which is available to authorized users.

✉ Sergio Romero
sromero@phhp.ufl.edu

¹ Department of Occupational Therapy, University of Florida, PO BOX 100164, Gainesville, FL 32610, USA

² Center of Innovation on Disability and Rehabilitation Research (CINDRR), Department of Veterans Affairs, 101 SE 2nd Place Ste 104, Gainesville, FL 32601, USA

³ Division of Occupational Therapy, Medical University of South Carolina, 151-B Rutledge Avenue, MSC 962, Charleston, SC 29425, USA

⁴ Division of Gerontology, Department of Epidemiology & Public Health, University of Maryland School of Medicine, Rm 213, Howard Hall, 660 west redwood, Baltimore, MD 21201, USA

⁵ Parkinson's Disease and Movement Disorders, Department of Neurology, University of Maryland School of Medicine, 110 S. Paca Street, Rm 3-S-127, Baltimore, MD 21201, USA

(e.g., physical ability, behavioral dysfunction, depression, and anxiety) and disability [4–8]. For example, Meredith et al. [7] reported that pain self-efficacy was more predictive of patients' disability than pain intensity. As such, self-efficacy is currently used to tailor self-management programs for managing health conditions (e.g., via goal setting, coping, or problem-solving skills) [9, 10].

It is important to note that self-efficacy is entirely behavior- and task-related [11]. For instance, an individual can have high self-efficacy for getting out of bed and low self-efficacy for climbing Mount Everest [3]. This is because the range of skills required for different behaviors and tasks vary by context [1]. Sometimes, self-efficacy is evaluated at a global level [12], or even with a single item [13], but these global judgements of self-efficacy can overlook variations across individuals' behavioral complexity [1]. Inferences from general self-efficacy measures have less explanatory and predictive power [1]. Thus, when measuring individuals' self-efficacy, we need to consider assessing it with respect to a specific behavior or task.

The patient-reported outcome measurement information system self-efficacy measure for managing chronic conditions (PROMIS-SE) is a behavior-specific self-efficacy measure for managing chronic conditions. Funded by the National Institute of Health (NIH), PROMIS-SE is a publicly available self-reported measure that is efficient and valid for individuals with chronic conditions [14]. Various qualitative methodologies (i.e., Delphi, focus groups, and cognitive interviews) were used to determine the PROMIS-SE's five domains. The PROMIS-SE's five domains measure respondents' self-efficacy for managing (1) daily activities; (2) emotions; (3) medications and treatments; (4) social interactions; and (5) symptoms [14, 15]. The PROMIS-SE estimates respondents' self-efficacy for each domain [14].

It is reasonable to expect patient's responses across PROMIS-SE domains to be influenced by multiple domains. That is, a therapeutic program aiming to enhance patient self-efficacy for managing emotions might improve not only the patient's self-efficacy levels for managing emotions, but also their self-efficacy levels for managing social interactions. This would indicate that perhaps these domains are not as unidimensional as previously assumed [14] and that other psychometric models (i.e., multidimensional models) might be more appropriate for calibrating these domains. It has been argued by some investigators that this might be the case for many of the existing patient-reported outcomes [16–18]. This multidimensionality of a construct (e.g., self-efficacy for managing chronic conditions) indicates that the level of one domain (e.g., self-efficacy for managing emotions) can also provide information about the level of another domain (e.g., self-efficacy for managing social interaction). Thus, simultaneous examination of PROMIS-SE domains might help clinicians better understand their patients' self-efficacy.

Patients' self-efficacy profiles can be created based on interactions among domains. Profiles can be used by clinicians to develop targeted interventions that might include multiple strategies across domains. For instance, in education, student mathematics profiles for the division of fractions can help teachers design individually tailored instructional math programs by identifying target skills for improvement (e.g., ability to convert mixed numbers, multiply fractions) [19].

However, the current PROMIS-SE provides individual scores for each domain with no possibility of obtaining a global SE profile for managing chronic conditions. PROMIS-SE measures that are not in the same metric (i.e., they are individually calibrated) do not account for the potential influences of interrelated domains. Our previous study reported high correlations among all PROMIS-SE domains, suggesting individuals' responses to a particular domain (e.g., managing symptoms) are heavily associated with their responses to another domain (e.g., managing emotions and social interactions, $r > 0.75$) [14]. While some advantages exist for treating domains as independent factors (less computational demands, smaller sample size requirement, and simpler outcome interpretation), when the correlations are strong among domains, incorporating interactions across domains can result in more accurate estimations [20, 21], in turn, leading to better therapeutic evaluation and planning.

Multidimensional models estimate respondents' abilities or functions based on the premise of having more than one conceptually and empirically defined latent construct that influences their responses to items on the measure; thus, correlations among domains can be integrated into measure estimations. In the context of PROMIS-SE, these models would allow measures to account for inter-domain correlations in the calibrations while preserving self-efficacy's behavior-specific characteristic (i.e., obtaining scores for each domain). Applying a multidimensional model to our PROMIS-SE patient data would enable users to (1) obtain simultaneous self-efficacy domain estimates that account for domain correlations; and (2) create a common metric for meaningful comparisons across domains. However, in order to apply a multidimensional model to a measure, one must first evaluate the assumption that multiple dimensions (i.e., factors; latent traits) underlie the item response data matrix. This can be evaluated by fitting a multidimensional model that evaluates model fit to the data.

Therefore, the purpose of the study is to (1) investigate current domain metric discrepancies in PROMIS-SE; (2) test the PROMIS-SE's multidimensional model fit; and (3) examine PROMIS-SE domain and item psychometric properties under a multidimensional framework. This study is part of a larger study that employs a multidimensional model to develop patient self-efficacy profiles concurrently across domains. Our two subsequent manuscripts

use patient responses to demonstrate the feasibility of using a multidimensional model to estimate patient self-efficacy for managing chronic conditions.

Methods

Data collection

A total of 1087 patients were recruited from two sites. Patients were recruited from a clinical practice at the University of Maryland Neurology Ambulatory Center ($n = 837$), and a national online recruitment company $n = 250$; (see op4g.com for more detail). Patients recruited from the University of Maryland Neurology Ambulatory Center had chronic neurologic conditions, whereas patients recruited from Op4G had general chronic conditions. Recruited patient participation was from April 2013 to April 2014 at the University of Maryland Neurology Ambulatory Center and from August to September 2013 from online recruitment.

Inclusion criteria for participants at the University of Maryland Neurology Ambulatory Center (treated by neurologists) were the following: 18 years of age or older, resided in the community, and diagnosed with one of the following chronic conditions: epilepsy, multiple sclerosis, Parkinson's disease, peripheral neuropathy, and stroke. Patients with cognitive impairment (i.e., scored below 20 on the Montreal Cognitive Assessment), inability to give informed consent, severe or unstable medical conditions, pregnant women, prisoners, and institutionalized patients were excluded from the sample at the University of Maryland Neurology Ambulatory Center.

For online recruitment, participants were randomly selected from approximately 250,000 Op4G subjects. The selected subjects completed web-based surveys using their personal devices. Inclusion criteria for the online sample were (1) 18 years of age or older, (2) community residence, and (3) having one of the following general chronic conditions: coronary artery disease, heart failure or congestive heart failure, heart attack (myocardial infarction), stroke or transient ischemic attack (TIA), liver disease, hepatitis, or cirrhosis, kidney disease, arthritis or rheumatism, asthma, chronic obstructive pulmonary disease (COPD), chronic bronchitis or emphysema, migraines or severe headaches, diabetes or high blood sugar or sugar in your urine, cancer (other than non-melanoma skin cancer), HIV or AIDS, spinal cord injury, multiple sclerosis, Parkinson's disease, neuropathy, or epilepsy. Participants were eligible for the study as long as they had at least one of the conditions listed above; additional health conditions (other than those listed) were allowed.

PROMIS-SE

The PROMIS-SE has three test forms (Computerized Adaptive Test (CAT), eight-item and four-item short forms) and five behavioral domains (daily activities (DA), Emotions (EM), medications and treatments (MT), social interactions (SS), and Symptoms (SX)). Test forms were created using domain-specific item banks. The five item banks included a total of 137 items (DA-35 items, EM-25 items, MT-26 items, SS-23 items, and SX-28 items). With CAT, individual item responses determine the selection of subsequent items from the item bank. Respondents are required to answer a minimum number of four and a maximum number of 12 items to obtain scores for each PROMIS-SE domain. The CAT stops when it reaches a standard error (of an estimated score) below 0.3 or the maximum number of administered items (i.e., 12). Short form items were selected based on item discrimination, item difficulty, and local independence from each domain item bank. The structural model for the original PROMIS-SE is illustrated in Fig. 1.

Our previous article reported that patient scores from two PROMIS-SE short forms were highly correlated with results using item banks ($r > 0.85$ for four-item and $r > 0.90$ for eight-item forms) [14]. Lists of items are provided in Online Appendix. Our previous article also reported that none of the items in our five item banks had sufficient differential item functioning (McFadden's pseudo $R^2 > 10\%$) to bias item estimates by age (under 65/over 65 years old), gender (male/female), race (white/non-white), or data source (the University of Maryland Neurology Ambulatory Center vs Op4G) [14].

All final patient scores are estimated based on the Item Response Theory (IRT) Graded Response Model (GRM). GRM links items to a latent construct. In our analysis, GRM links items to each PROMIS-SE domain (i.e., DA, EM, MT, SS, or SX). All calibrations were conducted separately for each domain, and all results were provided as T-scores (centered at 50 with a standard deviation of ten). These T-scores are standardized based on the distribution of our clinical sample and will be referred to as T_{clin} -scores hereafter. Higher scores indicate higher levels of self-efficacy for the particular domain. CAT directly provides users with T_{clin} -scores, but for short forms, scoring conversion tables should be used to convert the summated raw scores to T_{clin} -scores. Scoring tables are provided in the PROMIS website (http://www.healthmeasures.net/images/promis/manuals/PROMIS_Self_Efficacy_Scoring_Manual.pdf).

Statistical data analysis

For descriptive statistics, PROMIS-SE's T_{clin} -score ranges and distributions were examined to identify domain discrepancies across item banks and four-item and eight-item short

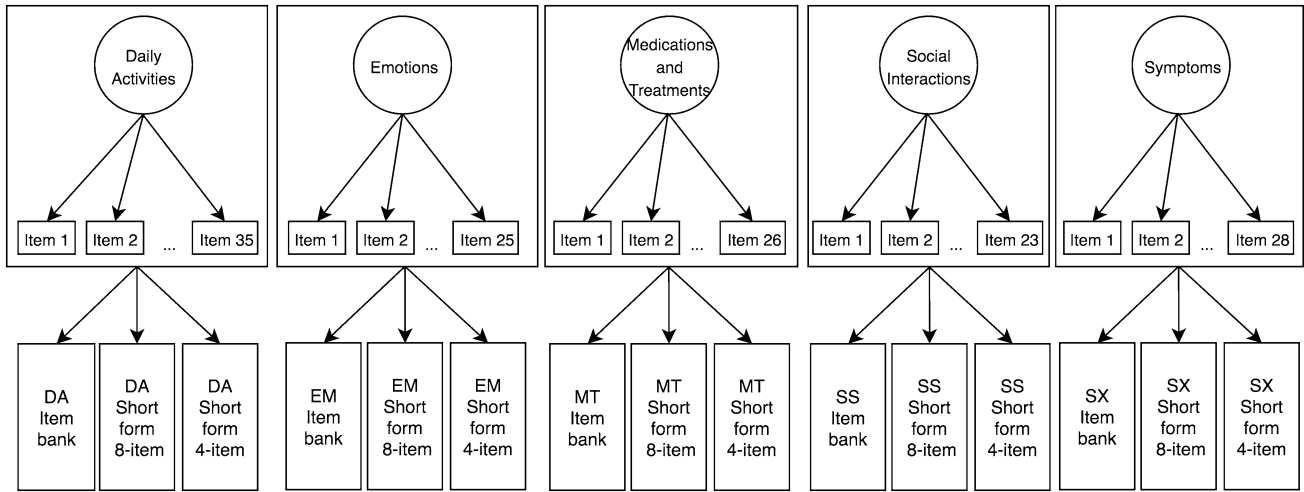


Fig. 1 PROMIS structural models

forms. In addition, a frequency analysis of item ratings was conducted to compare percentages for each item rating (i.e., total response numbers for each item rating/total response number) to investigate response patterns across PROMIS-SE domains.

The multidimensionality of the PROMIS-SE was tested using confirmatory factor analysis (CFA) for the item banks and the four-item and eight-item short forms (see Fig. 2) with the following criteria for fit-statistics: RMSEA (<0.06 good, <0.08 acceptable), CFI and TLI (>0.95 good, >0.90 acceptable), and SRMR (<0.05 good, <0.08 acceptable) [22]. Weighted least square mean and variance (WLSMV) was used as the estimator to accommodate the Likert responses of PROMIS-SE, and to maximize the efficiency of the measure, we used pairwise deletion. We also examined domain correlations, item factor loadings, and R-squares under a multidimensional model for PROMIS-SE item banks. For statistical analyses, R version 3.4.3, R studio version 1.1.414, and R packages (lavaan, dplyr, tidyr, and ggplot2) were used.

Results

Demographics

On average, participants had 3.8 chronic conditions. The top five commonly shared chronic conditions were hypertension (40%), arthritis or rheumatism (33%), depression (33%), anxiety (29%), and migraines or severe headaches (26%). Table 1 provides the prevalence for 25 chronic conditions. The mean age of the sample population was 53.6 years old (SD = 14.7). The proportion of gender was 42.1% for male and 55.1% for female participants (2.8% missing data). Please see Hong et al. [15] and Gruber-Baldini et al. [14] for additional demographic information.

Descriptive statistics

Our results indicate that the PROMIS-SE T_{clin} -score ranges had inconsistent patterns across the five domains. While

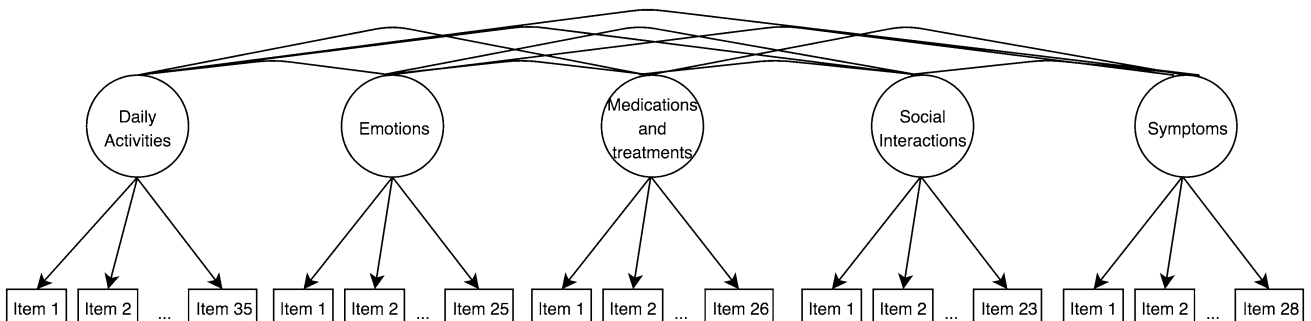


Fig. 2 Hypothesized PROMIS structural models

Table 1 Chronic conditions

Chronic conditions	Count (%)
High blood pressure (hypertension)	437 (40.2)
Arthritis or rheumatism	362 (33.3)
Depression	358 (32.9)
Anxiety	314 (28.9)
Migraines or severe headaches	284 (26.1)
Neuropathy	258 (23.7)
Sleep disorder	232 (21.3)
Stroke or transient ischemic attack (TIA)	201 (18.5)
Epilepsy	193 (17.8)
Diabetes or high blood sugar or sugar in your urine	187 (17.2)
Multiple sclerosis	183 (16.8)
Asthma	181 (16.7)
Parkinson's disease	177 (16.3)
Cancer (other than non-melanoma skin cancer)	124 (11.4)
Osteoporosis	118 (10.9)
Chest pain (angina)	99 (9.1)
Hardening of the arteries (coronary artery disease)	61 (5.6)
Chronic lung disease (COPD), chronic bronchitis or emphysema	59 (5.4)
Heart attack (myocardial infarction)	52 (4.8)
Spinal cord injury	50 (4.6)
Kidney disease	46 (4.2)
Alcohol or drug problem	38 (3.5)
Heart failure or congestive heart failure	38 (3.5)
Liver disease, hepatitis, or cirrhosis	37 (3.4)
HIV or AIDS	10 (0.9)

Table 2 T_{clin} -score ranges of PROMIS-SE domains

Domains	Item bank	8-item	4-item
Daily activities	50 (16–65)	38 (24–61)	34 (26–59)
Emotions	51 (19–69)	43 (23–65)	39 (25–63)
Medications and treatments	51 (13–63)	43 (19–61)	37 (22–58)
Social interactions	52 (14–65)	41 (20–60)	36 (23–58)
Symptoms	52 (18–69)	42 (23–64)	38 (25–62)

() indicates the upper and lower limits of T_{clin} -scores

PROMIS-SE EM had the highest (least negatively skewed) T_{clin} -scores (69, 65, and 63), MT had the lowest (most negatively skewed) T_{clin} -scores (13, 19, and 22) for item banks, eight-item short forms, and four-item short forms, respectively. Overall, T_{clin} -scores ranged from 13 to 69 for item banks, from 19 to 65 for eight-item short forms, and from 22 to 63 for four-item short forms. The average ranges of T_{clin} -scores were 51.2, 41.4, and 36.8 for item banks, eight-item short forms, and four-item short forms. Table 2 provides PROMIS-SE T_{clin} -score ranges and Fig. 3 illustrates T_{clin} -score distributions for PROMIS-SE domains.

Frequency analyses identified that patients were more inclined to report higher self-efficacy ratings for DA, MT, and SS as compared to EM and SX. The most frequently reported rating was “very confident” for all domains, and for three domains (DA, MT and SS), “very confident” had higher than 50% of the total responses. The percentage difference between the most and the second most frequently reported ratings (“very confident” and “quite confident”) were 48.24%, 45.30%, and 37.10% for DA, MT, and SS, whereas they were 6.76% and 14.19% for EM and SX, respectively. Figure 4 demonstrates the response frequencies for each rating for each PROMIS-SE domain.

Confirmatory factor analysis

Out of 148,919 (1087 respondents with 137 items) responses, a total of 147,859 (1,060 missing, <0.01%) responses were used for the analysis. CFA confirmed that the multidimensional model adequately fit item banks and two short forms. Corresponding fit indices for item banks, eight-item forms, and four-item forms were RMSEA (0.045, 0.069, and 0.070), CFI (0.911, 0.952, and 0.976), TLI (0.909, 0.949, and 0.972), and SRMR (0.061, 0.051, and 0.037), respectively. Overall, all domains also demonstrated high correlations under the multidimensional model ($r > 0.652$). PROMIS-SE EM and SX domains showed the highest correlations among all bivariate combinations ($r = 0.788$), closely followed by SX and SS ($r = 0.783$) and MT and SS ($r = 0.782$). Please see Table 3 for additional domain correlations. Moreover, all items highly or moderately loaded on their respective domains ($\lambda > 0.675$), and equal or larger than 45.6% of the variance of all items was explained by the model. Online Appendix PROMIS-SE Item-Level Information provides additional item-level information.

Discussion

This study confirmed that the T_{clin} -score ranges and distributions varied across the five PROMIS-SE domains and that a multidimensional model adequately fit the PROMIS-SE. Differences in T_{clin} -score ranges and distributions verified our notion that T_{clin} -scores from each PROMIS-SE domain should not be compared. Also, adequate fit-statistics for the multidimensional model indicated that applying a multidimensional model is viable and recommended for estimating multiple domains of PROMIS-SE.

Some might assume that T_{clin} -scores can be compared across PROMIS-SE domains because T_{clin} -scores are standardized scores. Standardized scores represent an individual's percentage with respect to a reference population (in our case, individuals with chronic conditions). The same T_{clin} -score, indeed, indicates the same percentage.

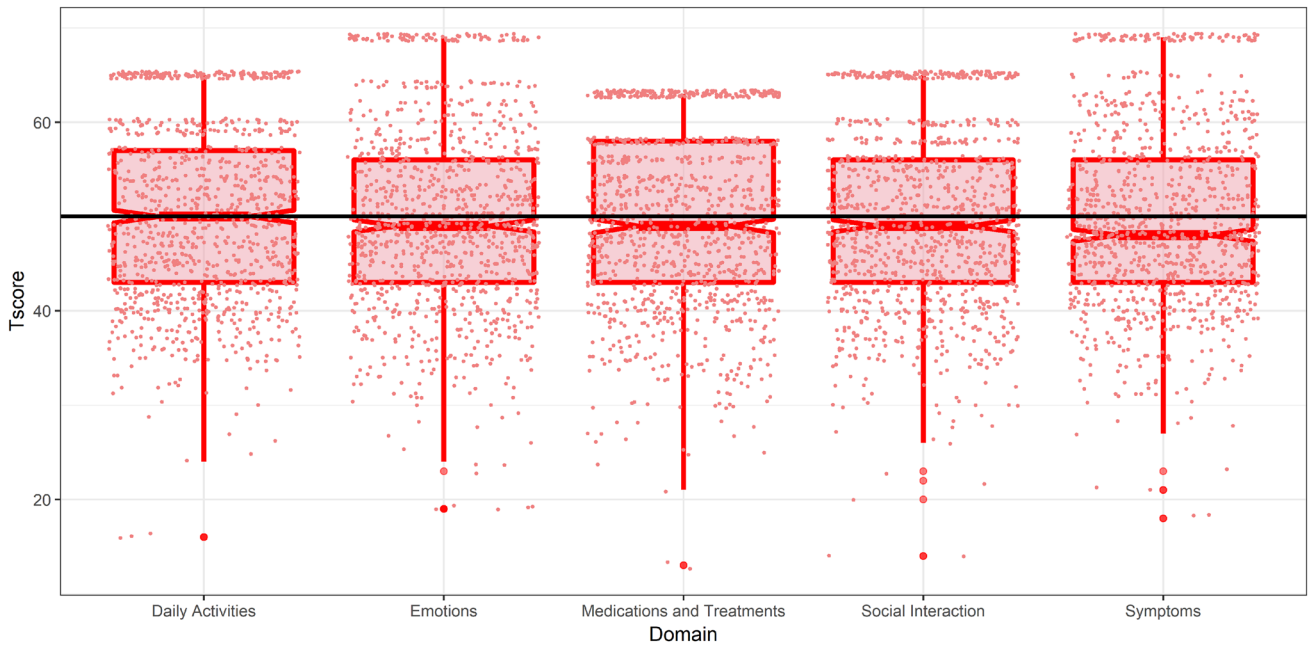


Fig. 3 T_{clin} -score distributions of PROMIS-SE item banks (box plot)

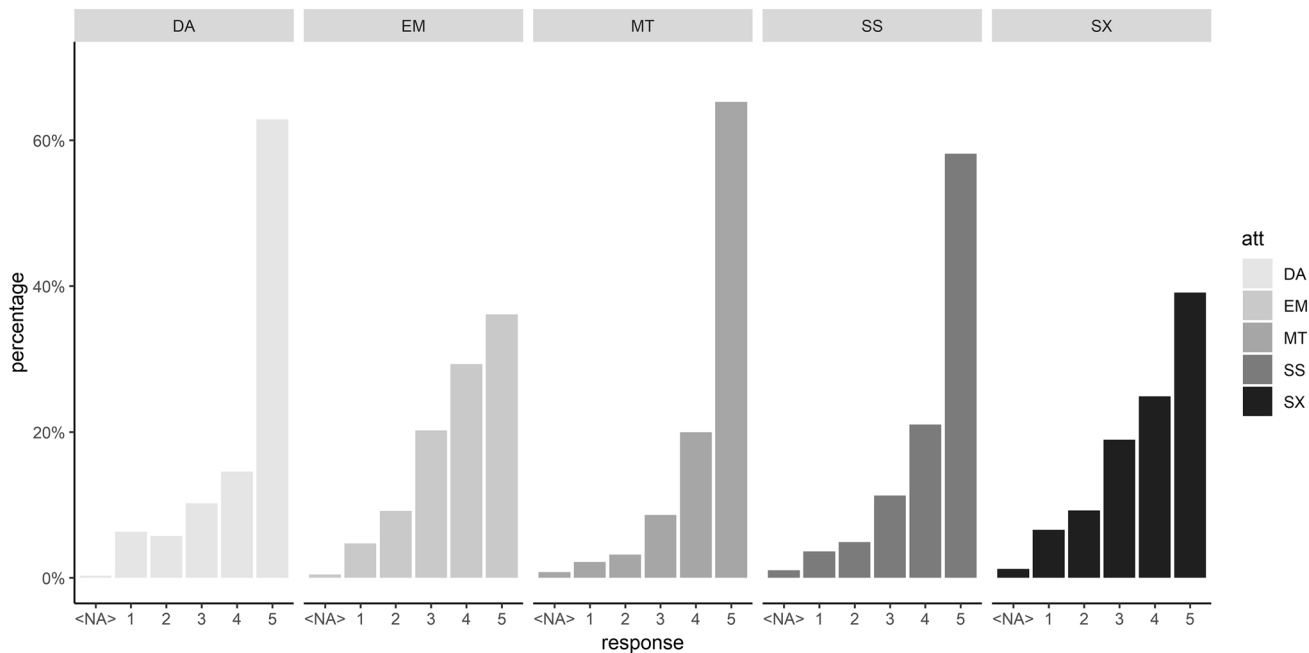


Fig. 4 Histogram of ratings for PROMIS-SE domains

However, each domain T_{clin} -score was calibrated on a different T_{clin} -score distribution, resulting in varied T_{clin} -score ranges, which may cause inaccurate interpretations of patient self-efficacy. For example, the maximum T_{clin} -scores for the MT and EM domains of the four-item short forms were 58 and 63, sequentially. Different maximum T_{clin} -scores correspond to different portions of the

population. For instance, responding “very confident” to all items on MT (21.19%) and EM (9.68%) may indicate that, overall, EM is more difficult than MT for this population. When a patient receives a T_{clin} -score of 58 on the MT and EM domains of the four-item short forms, some might interpret that this patient has the same level of self-efficacy for MT and EM. However, whether this patient has the

Table 3 Domain correlations under a multidimensional model

Domains	Daily activities	Emotions	Medication ^a	Social ^b	Symptoms
Daily activity		0.652	0.729	0.652	0.723
Emotion	0.652		0.709	0.755	0.788
Medication	0.729	0.709		0.782	0.754
Social	0.652	0.755	0.782		0.783
Symptom	0.723	0.788	0.754	0.783	

^aIndicates medications and treatments

^bIndicates social interactions

same level of self-efficacy for MT and EM is questionable, since the difficulties of the two domains may not remain the same. We suggest that T_{clin} -score interpretation should be done with caution or that practitioners be trained extensively in how to interpret these norm-referenced scores, especially when multiple domains are used together (i.e., the T_{clin} -scores can be compared to see how far away a person lies from the average score in standard deviation units of that domain scale, but this information should not be compared across domains as the average score and standard deviation of scores can vary across domains).

Multidimensional psychometric models that take into account inter-domain dependencies can be used to avoid confusion when comparing T_{clin} -score distributions across domains. Our results indicate the factor structure for the different forms of the PROMIS-SE (the item banks and two short forms (four-item and eight-item)) can be considered multidimensional. This was expected based on the high correlations reported in our original study [14]. Of note, while our multidimensional model showed satisfactory fit-statistics for all four indices (CFI, TLI, RMSEA, and SRMR), in our original study (treating domains as independent constructs) not all fit indices were adequate [14]. Using a model with better fit-statistics is desired because it allows for less biased measurement estimates. Fayers [23] reported that multidimensional models are more useful when domains have high correlations (as is the case with the PROMIS-SE) by allowing users to identify unusual scoring patterns that are often hidden under a unidimensional framework.

Along these lines, health-related items typically represent multiple domains and those domains are often highly correlated. Chang and Reeve [16] argue that almost all health-related outcomes are multidimensional. Multidimensional IRT (MIRT) models have demonstrated enhanced measurement precision and efficiency for various health-related outcomes including quality of life, pediatric functional skills, depression, anxiety, and global physical health [24–27]. In addition to MIRT, other multidimensional models, such as the bifactor model, locally dependent unidimensional models, and diagnostic classification models (DCMs), have also been introduced in education and psychology [28, 29].

A MIRT model can provide a patient PROMIS-SE profile and an overall set of scores (calculated concurrently across domains) (e.g., a profile of 59-50-58-50-50 for DA, EM, MT, SS, and SX, respectively). Similarly, multidimensional categorical models such as DCMs can provide categorical PROMIS-SE profiles (e.g., high-low-high-low-low self-efficacy) for five domains (DA, EM, MT, SS, and SX, respectively), where high/low or other appropriate thresholds indicate the level of proficiency. These patient profiles can be used to examine common self-efficacy patterns across different demographic groups, such as diagnosis, gender, age, and others. Clinical decisions for certain interventions could be inferred from patient profile results.

With rapid advances in computational power and the development of new software applications, these more complicated models, such as MIRT or DCMs, are expected to be more frequently applied to health-related constructs. Future studies are encouraged to apply these more complicated models and investigate practitioner perceptions of the different scoring results.

Our study confirmed the multidimensional nature of the PROMIS-SE. Important information about the relationship among domains that might lead to improvements in self-efficacy can be obtained when applying a multidimensional model. Clinicians can use this information to customize treatment and provide interventions that target the most appropriate items across domains to improve a patients' self-efficacy to manage their chronic condition. Other patient-reported outcome measures might face similar challenges. For example, we might expect measures such as the PROMIS depression and anxiety scales to be highly correlated. We suggest future studies to apply multidimensional models (e.g., Multidimensional IRT, Diagnostic Classification Modeling) to other health-related measures and further investigate the clinical utility of the measures from the clinicians' and patients' perspectives.

Funding The study was funded by the National Institutes of Health, *Development and Validation of a Self-Efficacy Item Bank* (Grant No. 1U01AR057967-01, Lisa Shulman (PI)). The presented results and conclusions in this paper are from the authors; the findings from this study are independent from the funding source. The fully developed

measures of self-efficacy for self-management of chronic diseases can be found at <http://www.healthmeasures.net/>.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval This study was approved by the Institutional review boards (IRB) of the Medical University of South Carolina (#Pro00033397), the University of Florida (#261–2010), and the University of Maryland (#HP-000432550). All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent All participants were treated and reimbursed as consultants. Therefore, informed consent was not required.

References

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W H Freeman and Company.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs: Prentice-Hall, Inc.
- Strecher, V. J., McEvoy DeVellis, B., Becker, M. H., & Rosenstock, I. M. (1986). The role of self-efficacy in achieving health behavior change. *Health Education Quarterly*, *13*(1), 73–92.
- McAuley, E., Konopack, J. F., Motl, R. W., Morris, K. S., Doerksen, S. E., & Rosengren, K. R. (2006). Physical activity and quality of life in older adults: Influence of health status and self-efficacy. *Annals of Behavioral Medicine*, *31*(1), 99–103.
- Tsay, S.-L., & Healstead, M. (2002). Self-care self-efficacy, depression, and quality of life among patients receiving hemodialysis in Taiwan. *International Journal of Nursing Studies*, *39*(3), 245–251.
- Muris, P. (2002). Relationships between self-efficacy and symptoms of anxiety disorders and depression in a normal adolescent sample. *Personality and Individual Differences*, *32*(2), 337–348.
- Meredith, P., Strong, J., & Feeney, J. A. (2006). Adult attachment, anxiety, and pain self-efficacy as predictors of pain intensity and disability. *Pain*, *123*(1), 146–154.
- Beckham, J. C., Burker, E. J., Burker, E. J., Feldman, M. E., & Costakis, M. J. (1997). Self-efficacy and adjustment in cancer patients: A preliminary report. *Behavioral Medicine*, *23*(3), 138–142.
- Siegert, R. J., & Levack, W. M. (2014). *Rehabilitation goal setting: Theory, practice and evidence*. Boca Raton: CRC Press.
- Newman, S., Steed, E., & Mulligan, K. (2008). *Chronic physical illness: Self-management and behavioural interventions*. New York: McGraw-Hill Education (UK).
- Lenz, E. R., & Shortridge-Baggett, L. M. (2002). *Self-efficacy in nursing: Research and measurement perspectives*. New York: Springer Publishing Company.
- Lorig, K., Chastain, R. L., Ung, E., Shoor, S., & Holman, H. R. (1989). Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, *32*(1), 37–44.
- Rebok, G. W., & Balcerak, L. J. (1989). Memory self-efficacy and performance differences in young and old adults: The effect of mnemonic training. *Developmental Psychology*, *25*(5), 714.
- Gruber-Baldini, A. L., Velozo, C., Romero, S., & Shulman, L. M. (2017). Validation of the PROMIS® measures of self-efficacy for managing chronic conditions. *Quality of Life Research*, *26*(7), 1915–1924.
- Hong, I., Velozo, C. A., Li, C.-Y., Romero, S., Gruber-Baldini, A. L., & Shulman, L. M. (2016). Assessment of the psychometrics of a PROMIS item bank: self-efficacy for managing daily activities. *Quality of Life Research*, *25*(9), 2221–2232.
- Chang, C.-H., & Reeve, B. B. (2005). Item response theory and its applications to patient-reported outcomes measurement. *Evaluation & the Health Professions*, *28*(3), 264–282.
- Pincus, T., Swearingen, C., & Wolfe, F. (1999). Toward a multidimensional Health Assessment Questionnaire (MDHAQ): Assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. *Arthritis & Rheumatology*, *42*(10), 2220–2230.
- Gold, J. I., Mahrer, N. E., Yee, J., & Palermo, T. M. (2009). Pain, fatigue and health-related quality of life in children and adolescents with chronic pain. *The Clinical Journal of Pain*, *25*(5), 407.
- Ketterlin-Geller, L. R., & Yovanoff, P. (2009). Diagnostic assessments in mathematics to support instructional decision making. *Practical Assessment, Research & Evaluation*, *14*(16), 1–11.
- Cheng, Y.-Y., Wang, W.-C., & Ho, Y.-H. (2009). Multidimensional Rasch analysis of a psychological test with multiple subtests: A statistical solution for the bandwidth—fidelity dilemma. *Educational and Psychological Measurement*, *69*(3), 369–388.
- Zhang, J. (2004). Comparison of unidimensional and multidimensional approaches to IRT parameter estimation. *ETS Research Report Series*, *2004*(2), i-40.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.
- Fayers, P. M. (2007). Applying item response theory and computer adaptive testing: The challenges for health outcomes assessment. *Quality of Life Research*, *16*(S1), 187–194.
- Petersen, M. A., Groenvold, M., Aaronson, N., Fayers, P., Sprangers, M., Bjorner, J. B., & for the European Organisation for Research and Treatment of Cancer Quality of Life Group (2006). Multidimensional computerized adaptive testing of the EORTC QLQ-C30: Basic developments and evaluations. *Quality of Life Research*, *15*(3), 315–329.
- Haley, S. M., Ni, P., Ludlow, L. H., & Fragala-Pinkham, M. A. (2006). Measurement precision and efficiency of multidimensional computer adaptive testing of physical functioning using the pediatric evaluation of disability inventory. *Archives of Physical Medicine and Rehabilitation*, *87*(9), 1223–1229.
- Haley, S. M., Ni, P., Dumas, H. M., Fragala-Pinkham, M. A., Hambleton, R. K., Montpetit, K., Bilodeau, N., Gorton, G. E., Watson, K., Tucker, C. A. (2009). Measuring global physical health in children with cerebral palsy: Illustration of a multidimensional bi-factor model and computerized adaptive testing. *Quality of Life Research*, *18*(3), 359–370.
- Bass, M., Morris, S., & Neapolitan, R. (2015). Utilizing multidimensional computer adaptive testing to mitigate burden with patient reported outcomes. In *AMIA Annual Symposium Proceedings* (Vol. 2015, p. 320). American Medical Informatics Association.
- Ip, E. H. (2010). Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response

- models. *British Journal of Mathematical and Statistical Psychology*, 63(2), 395–416.
29. Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. New York: Guilford Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.