Calibration and initial validation of a general self-efficacy item bank and short form for the NIH PROMIS®

John M. Salsman[1,2] · Benjamin D. Schalet[3] · Thomas V. Merluzzi[4] · Crystal L. Park[5] · Elizabeth A. Hahn[3,6] · Mallory A. Snyder[7] · David Cella[3,6]

Abstract

Aims General self-efficacy is associated with adaptive coping and positive health outcomes. The Patient-Reported Outcomes Measurement Information System (PROMIS®) has developed self-efficacy item banks for managing chronic conditions, but lacks a general self-efficacy measure. We sought to refine and validate an item-response theory (IRT)-based measure of general self-efficacy for PROMIS®.

Methods Ten items were modified from the NIH Toolbox® Self-Efficacy Item Bank by creating "confidence" response options, and administered to a general population sample ($n = 1000$) with the Toolbox® Self-Efficacy Item Bank, Life Orientation Test-Revised (LOT-R), and Generalized Expectancy for Success Scale (GESS). We split the sample in half for exploratory factor analysis (EFA) followed by confirmatory factor analysis (CFA). IRT analyses included evaluation of differential item functioning (DIF).

Results Participants had a mean age of 47.8 years and 50.3% were male. EFA showed all items loaded onto one dominant factor and CFA yielded a good fitting model for a general self-efficacy bank with confidence response options (CFI = 0.987, TLI = 0.984, RMSEA = 0.090). Items showed no evidence of DIF by gender, age, education, or race. Internal consistency reliability was $\alpha = .94$ and .88 for a new 10-item general self-efficacy bank and 4-item short form, respectively. The new bank was correlated with the LOT-R ($r = .58$), the GESS ($r = .55$), and the Toolbox® Self-Efficacy Item Bank ($r = .87$).

Conclusions The PROMIS® General Self-Efficacy measure demonstrated sufficient unidimensionality and displayed good internal consistency reliability, model fit, and convergent validity. Further psychometric testing of the PROMIS® General Self-Efficacy Item Bank and Short Form can evaluate its utility in people with chronic health conditions.

Keywords Oncology · Patient-reported outcomes · Item-response theory · Self-efficacy · Psychological adaptation · Well-being

Introduction

Self-efficacy refers to an individual's belief in his or her capacity to perform a particular behavior or set of behaviors. Strictly speaking, self-efficacy reflects confidence in one's

✉ John M. Salsman
jsalsman@wakehealth.edu

1 Department of Social Sciences and Health Policy, Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA

2 Wake Forest Baptist Comprehensive Cancer Center, Medical Center Boulevard, Winston-Salem, NC 27157, USA

3 Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

4 Department of Psychology, University of Notre Dame, Notre Dame, IN, USA

5 Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA

6 The Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Chicago, IL, USA

7 University Research Administration, The University of Chicago, Chicago, IL, USA

ability to exert control over one's own motivation and behavior regardless of the outcome [1]. Theoretically informed descriptions of self-efficacy emphasize the role of autonomy, self-determination, mastery, and self-regulation [1–3]. Self-efficacy is associated with adaptive coping strategies and more positive health outcomes. It can be a catalyst for better self-management and improved outcomes in chronic health conditions such as rheumatoid arthritis, cardiac complaints, stroke, and cancer [4–8]. Moreover, because self-efficacy reflects a sense of control and personal agency [1, 2], it is often an independently important and valuable outcome within the context of patient-centered care and is a key component of models that are used to predict health-related intentions and behaviors [9].

Self-efficacy can be conceptualized as having both a general (global) component and more narrowly defined, behavior-specific components. General self-efficacy may reflect a more stable, personality-driven construct of psychological hardiness, optimism, or resilience [10, 11]. It suggests a favorable attitude or positive set of expectations that emphasizes a problem-solving approach despite perceived obstacles or challenges. More narrowly defined aspects of self-efficacy may be focused on relatively discrete aspects of health behaviors (e.g., self-efficacy for exercise, self-efficacy for maintaining a healthy diet) [12, 13], coping skills (e.g., self-efficacy for managing emotions) [14], or symptom management (e.g., self-efficacy for managing fatigue) [15], among other important aspects of a person's experience.

The Patient-Reported Outcomes Measurement Information System (PROMIS) is the most comprehensive approach to standardizing assessment of health-related quality of life in acute and chronic health conditions [16]. By leveraging item-response theory (IRT), multiple measures of symptoms and functioning have been developed for flexible (computer adaptive tests), efficient (minimal burden on patients), and precise (reduced measurement error) assessments and have demonstrated clinical validity across diverse chronic conditions [17, 18]. Importantly, PROMIS has developed self-efficacy item banks and short forms for managing symptoms, daily activities, social interactions, medications and treatment, and emotions [19]. However, PROMIS is lacking in a comparable measure of global or general self-efficacy.

The NIH Toolbox® includes a measure of General Self-Efficacy that was adapted from the Generalized Self-Efficacy Scale [10] and subsequently tested and refined through similar IRT approaches as PROMIS [20]. The NIH Toolbox General Self-Efficacy Scale is a logical choice to complement the PROMIS context-specific item banks and short forms for self-efficacy, but the response options use a frequency (i.e., "never" to "very often") format. Self-efficacy theory would suggest that confidence response options (i.e., "I am not at all confident" to "I am very confident"), which focus on behavior expectancy as opposed to solely on

prior behavior, better reflect the underlying construct [21]. Moreover, the PROMIS domain-specific self-efficacy item banks and short forms use confidence response options, and patients in cognitive debriefing interviews preferred them over other options [22].

We sought to address these gaps by (1) refining a patient-reported outcome assessment tool of general self-efficacy for PROMIS and evaluating assumptions for IRT that are consistent with PROMIS Scientific Standards (e.g., unidimensionality and local independence) [23], (2) examining item-level properties to support computer adaptive testing (CAT) and evaluate possible differential item functioning (DIF), and (3) identifying a static short form and examining convergent validity of the newly developed PROMIS General Self-Efficacy Short Form and Item Bank.

Methods

Participants and procedures

We partnered with Opinions for Good (http://op4g.com/), an online research panel company, to recruit adult (ages 18 or older), English-speaking participants from the US general population. Liu et al. [24] have shown the representativeness of internet data is comparable to data from probability-based general population samples, and the internet is a low cost and efficient means of data collection that is widely accessible to diverse groups [25]. This study was approved by the Institutional Review Board of Northwestern University. All interested and eligible participants provided informed consent electronically.

To recruit study participants, Op4G sent emails to a random selection of panel members from their databases to invite them to enroll in the current study. We pre-specified target distributions for age/gender (minimum $n = 300$ in each of three age strata "18–39", "40–59", and "60–85" with a minimum $n = 120$ men and 120 women in each subgroup), race and ethnicity (minimum $n = 200$ participants who self-identify as Hispanic or Latino and minimum $n = 200$ participants who self-identify as Black or African American), and educational attainment (minimum $n = 400$ for high school graduate/GED or less and minimum $n = 400$ for some college or greater). Following a screening process to ensure eligibility, participants provided informed consent and then completed a demographic survey and other self-report measures (described below). To reduce the potential for order effects, all measures were administered in random, thematic blocks and the order of items within a bank were also randomized. Participants who completed questionnaires were eligible for incentive-based compensation and donations made to a charity of their choosing through Op4G. For calibration

and validation purposes, the newly modified items were administered to a general population sample ($n = 1000$).

Study Measures

NIH Toolbox Self-Efficacy Item Bank

The NIH Toolbox Self-Efficacy Item Bank is a 10-item, calibrated bank derived from the Generalized Self-Efficacy Scale [10] designed to assess a person's belief in his/her capacity to manage daily stressors and have control over meaningful events [20]. The NIH Toolbox Self-Efficacy Item Bank uses a Likert scale with frequency response options ("Never", "Almost Never", "Sometimes", "Fairly Often", and "Very Often"). Higher scores reflect greater general self-efficacy.

PROMIS General Self-Efficacy Item Bank

Informed by theory and qualitative input from patients and content experts [22], 10 items were modified from the NIH Toolbox Self-Efficacy Item Bank by creating new "confidence" response options that mirrored the same response options as the PROMIS® measures of Self-Efficacy for Managing Chronic Conditions ("I am not at all confident", "I am a little confident", "I am somewhat confident", "I am quite confident", "I am very confident") [19]. Higher scores reflect greater general self-efficacy.

Life Orientation Test-Revised (LOT-R)

The LOT-R is a self-report measure of optimism that consists of six items plus fillers [26]. Each item is rated on a 4-point Likert scale that ranges from "I agree a lot" to "I disagree a lot." Three of the items are framed positively (e.g., "In uncertain times I expect the best"), and three of the items are framed negatively and reverse-scored (e.g., "If something can go wrong for me it will"). Higher scores reflect greater optimism.

Generalized Expectancy for Success Scale-Short Form (GESS-SF)

The GESS-SF is a four-item, self-report measure used to evaluate participants' expectancies for future events [27, 28]. Sample items include "In the future I expect that I will experience many failures in my life" and "In the future I expect that I will be unable to accomplish my goals," rated on a 5-point Likert scale from "definitely not" to "definitely." The items were recoded so that higher scores on the GESS-SF represented higher expectancy for future-oriented goal attainment.

PROMIS Global-10

The PROMIS Global-10 is a 10-item short form that assesses general domains of health and functioning, including overall physical health, mental health, social health, pain, fatigue, and overall perceived quality of life [29]. We used the two, four-item summary scores for this project: Global Physical Health and Global Mental Health.

Statistical analysis

Analyses followed general guidelines used in the PROMIS item bank development [23, 30, 31] and were grouped into three stages: (1) testing assumptions for IRT modeling—unidimensionality and local independence of items, (2) estimating item parameters using IRT and evaluating items for DIF, and (3) selecting items for static short forms and examining preliminary validity. Final item inclusion/exclusion was decided by group consensus after reviewing analytic results and item content.

During the first stage, we examined items for sparse data within any rating scale response category (i.e., $n < 5$). Data were randomly divided into two datasets ($n = 500$ each), one for exploratory factor analysis (EFA) and the other for confirmatory factor analysis (CFA), using the psych package in R [32, 33] and MPlus 7.2 (Muthen and Muthen, Los Angeles, CA), respectively. EFA of the polychoric correlation matrix with oblique rotation was used to identify potential factors among items and CFA was used to confirm final factor structure. In the EFAs, eigenvalues > 1.0 and scree plots were used as criteria to estimate meaningful factors. Items with factor loadings < 0.4 were considered for exclusion. Next, we estimated the proportion of total variance attributable to a general factor with omega hierarchical (omega-h) using the **psych** package [33]. This method estimates omega-h from the general factor loadings derived from an EFA and a Schmid–Leiman transformation [34]. Values of .70 or higher suggest that the item set is sufficiently unidimensional [35].

In the CFAs we used the weighted least squares estimator and fit statistics to evaluate dimensionality of the item pool. We selected the commonly used indices for item banking as recommended by PROMIS Scientific Standards: comparative fit index (CFI), Tucker–Lewis index (TLI), and root mean square error of approximation (RMSEA). We used the following model fit indices as guidelines: RMSEA $< .08$, CFI $> .95$, TLI $> .95$. Residual correlations were used to identify locally independent item pairs ($< .10$) and to avoid potential secondary factors from locally dependent items [31].

In the second stage, the total sample ($n = 1000$) was used and items that met unidimensionality assumption were analyzed using Samejima's graded response model (GRM) [36]

as implemented in IRTPRO [37]. The GRM yields threshold (location) and slope parameters. Item threshold parameters locate items along the measured trait and show the coverage across the general self-efficacy continuum. The item slope parameter represents the discriminative ability of the items, with higher slope values indicating better ability to discriminate between adjoining values on the construct. Items displaying poor IRT fit (criterion: significant S-X^2 fit statistic, $p < 0.05$ [38, 39]) and poorly discriminating items (i.e., those with unacceptable IRT slopes; criterion: slope < 1) were candidates for exclusion at this stage. We used LORDIF to conduct DIF analyses on the basis of age ("18–39" vs. "40–59", "18–39" vs. "60–85", "40–59" vs. "60–85"), gender ("male" vs. "female"), education ("≤high school" vs. ">high school"), and race ("white" vs. "non-white," "black" vs. "non-black") for groups with a minimum of 200 participants per subgroup [35]. An item has significant DIF if it exhibits different measurement properties between subgroups, which is similar to "item bias." DIF exists when characteristics such as age, gender, education, or race, which may seem insignificant to the assessment of domains of interest, have an effect on measurement. Specifically, we tested for DIF using an ordinal logistic regression procedure [36] with χ^2 to detect items ($p < 0.01$), and McFadden pseudo $R^2 > 0.02$ as the threshold for substantial DIF [40]. Items that demonstrated DIF greater than $R^2 > 0.02$ were to be removed.

In the third and final stage, a fixed-length short form was determined by group discussion and consensus. Our team of psychometricians, content-expert consultants, and measurement scientists reviewed item content, threshold, and slopes for all general self-efficacy items in the calibrated bank to identify an optimal short form. Finally, the convergent validity of the PROMIS General Self-Efficacy Item Bank and Short Form was examined using bivariate Pearson correlations with comparable constructs. We hypothesized that the PROMIS General Self-Efficacy Item Bank and Short Form would demonstrate the largest correlations with the NIH Toolbox Self-Efficacy Item Bank but would also be significantly correlated with the LOT-R, GESS-SF, and the PROMIS Global Mental Health scores and less strongly correlated with the PROMIS Global Physical Health scores.

Results

Sample characteristics

Our sample comprised approximately equal numbers of young (ages 18 to 39), middle-aged (ages 40 to 59), and elderly (ages 60 to 85) adults. It was predominantly White (68.3%) but had good representation from racial and ethnic minorities. Approximately equal numbers of participants

had received a high school education or less compared to those who had some college education or greater. Additional demographic characteristics are shown in Table 1.

IRT assumptions

We examined frequencies for both versions of the General Self-Efficacy Scale to ensure adequate numbers of responses for each category for all items. Confidence response options captured by the new PROMIS General Self-Efficacy Item Bank resulted in scores with a lower mean and wider distribution ($M = 34.8$, SD $= 8.7$) relative to the NIH Toolbox Self-Efficacy Item Bank ($M = 37.3$, SD $= 7.2$). In addition, PROMIS scores were not as concentrated at the top compared to Toolbox (11% vs. 15% for the highest possible 5 scores), while more scores were close to the bottom (2% vs. 0.04% for the lowest possible 5 scores). See Fig. 1. In addition, the PROMIS items had slightly higher item-total correlations than did the NIH Toolbox items, $r = 0.68$ to 0.79 compared with $r = 0.61$ to 0.77, respectively.

In order to establish the relative unidimensionality of the PROMIS General Self-Efficacy Item Bank, we randomly split the sample into two halves ($n = 500$ each) and conducted an EFA followed by a CFA. We ran single factor EFA models, using the weighted least squares method, based on the polychoric correlation matrix. An examination of the scree plot suggested there was one dominant factor (Fig. 2) with all items loading on the primary factor and the PROMIS items accounting for more explained variance than the Toolbox items (69% to 59%). In addition, omega-h values for both PROMIS and Toolbox versions were very high, 0.87 and 0.93, respectively, suggesting the presence of a dominant general factor. Consequently, all 10 items from both measures were retained for the subsequent CFA.

We then conducted a CFA on the other half of the sample ($n = 500$), paying particular attention to model fit indices (RMSEA < .08, CFI > .95, TLI > .95) and residual correlation (< .10). Based on comparable (and acceptable) fit statistics for the confidence items (CFI $= 0.99$, TLI $= 0.98$, RMSEA $= 0.09$, 90% CI $= 0.07$ to 0.10, $\chi^2 = 177.87$, d.f. $= 35$, $p < .0001$) compared to the frequency items (CFI $= 0.99$, TLI $= 0.99$, RMSEA $= 0.08$, 90% CI $= 0.06$ to 0.09, $\chi^2 = 139.83$, d.f. $= 35$, $p < .0001$) and no elevated residual correlations (all < .10), we decided that the 10-item PROMIS General Self-Efficacy Item Bank with confidence response options was sufficiently unidimensional and free of local dependence.

Estimating item parameters and evaluating DIF

Once we established unidimensionality and local independence, the next step was to calibrate the new general self-efficacy bank using estimated IRT parameters from a GRM

Table 1 Demographic characteristics (n = 1000)

	47.8 (M)	16.2 (SD)
	n	%
Age		
18–39	341	34.1
40–59	335	33.5
60–85	324	32.4
Gender		
Female	497	49.7
Ethnicity		
Hispanic origin	199	19.9
Race		
White	683	68.3
Black/African American	200	20.0
Asian or Pacific Islander	41	4.1
Native American or Alaskan Native	37	3.7
Native Hawaiian or Other Pacific Islander	14	1.4
Other	50	5.0
Education		
Eighth grade or less	50	5.0
Some high school	114	11.4
High school grad/GED	334	33.4
Some college/technical degree/AA	217	21.7
College degree (BA/BS)	203	20.3
Graduate school	82	8.2
Country of origin		
US	897	89.7
Other	103	10.3
Relationship status		
Married/living with partner in committed relationship	548	54.8
Never married	267	26.7
Divorced/separated	154	15.4
Widowed	31	3.1
Times stayed in hospital overnight or longer in past 12 months		
0	672	67.2
1–2	206	20.6
3–4	54	5.4
5–6	31	3.1
7–10	22	2.2
11–50	13	1.3
51–75	2	0.2
Days that poor physical or mental health kept you from doing normal activities in past 30 days		
0	448	44.8
1–2	193	19.3
3–4	96	9.6
5–6	77	7.7
7–10	75	7.5
11–30	111	11.1
Comorbidities		
High blood pressure (hypertension)	392	39.2
Anxiety	277	27.7
Depression	270	27.0
Arthritis or rheumatism	266	26.6

Table 1 (continued)

|	47.8 (M)	16.2 (SD)
	n	%
Migraines or severe headaches	242	24.2
Sleep disorder	183	18.3
Asthma	175	17.5
Diabetes or high blood sugar or sugar in your urine	145	14.5
Chest pain (angina)	135	13.5
Alcohol or drug problem	72	7.2
Chronic lung disease (COPD), chronic bronchitis or emphysema	68	6.8
Heart attack (myocardial infarction)	64	6.4
Liver disease, hepatitis, or cirrhosis	60	6.0
Cancer (other than non-melanoma skin cancer)	60	6.0
Hardening of the arteries (coronary artery disease)	56	5.6
Heart failure or congestive heart failure	46	4.6
Stroke or transient ischemic attack (TIA)	42	4.2
Kidney disease	42	4.2
Spinal cord injury	42	4.2
Multiple sclerosis	36	3.6
HIV or AIDS	16	1.6
None of the above	4	0.4
Used/taken prescription medication in past 30 days		
Yes	646	64.6
No	340	34.0
Don't know	14	1.4
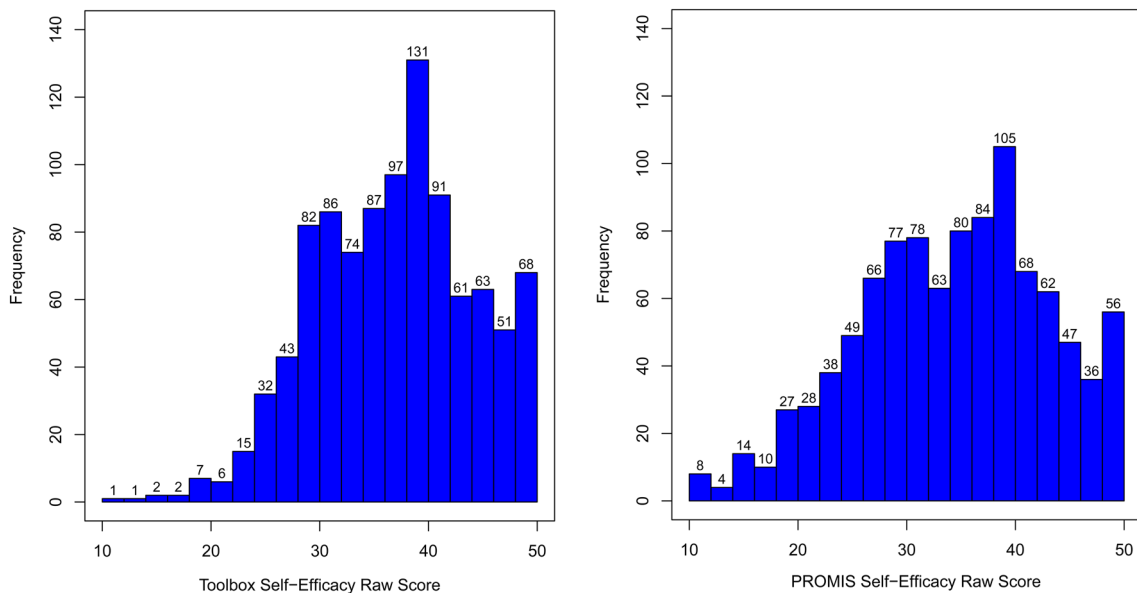
Fig. 1 Frequency distribution of raw scores

to inform item slope (discrimination) and threshold (location) parameters. All item slopes were > 1.0, which met our inclusion criteria with the average slope = 2.45. The location parameters ranged from − 2.94 to 1.29. However, four

items suggested a poor fit ($S\text{-}X^2 < .01$) and were candidates for exclusion ("I can manage to solve difficult problems if I try hard enough," "I can solve most problems if I try hard enough," "I stay calm when facing difficulties because I can

Fig. 2 Scree plots

NIH Toolbox® Self-Efficacy Item Bank

PROMIS® General Self-Efficacy Item Bank
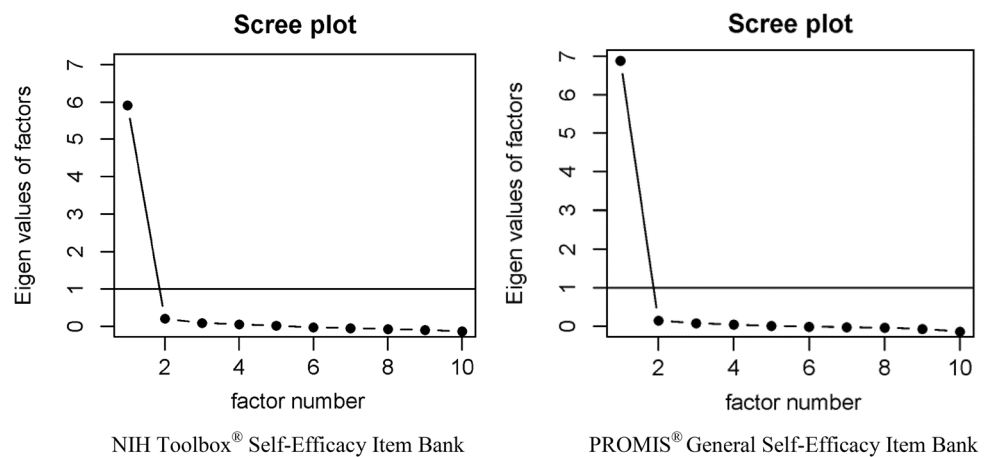
handle them," "If I am in trouble, I can think of a solution"). Investigating further, we scored these four items and the six fitting items separately using the IRT parameters from the model. The mean difference in resulting scores was 0.13 T-score points, suggesting minimal bias due to the poor fit. In addition, the general factor loadings from the Schmid–Leiman output for these 4 items were all higher than .80, suggesting little distortion due to specific or unique factor variance.

None of the items exceeded the McFadden pseudo R^2 threshold of 0.02 in any of the DIF comparisons, therefore showing no non-trivial effects for gender, age, education, and race on the latent trait of self-efficacy. Since all 10 items displayed good discrimination (i.e., slope parameters), no substantial DIF, showing minimal mean bias, and were derived from a commonly used legacy measure, the Generalized Self-Efficacy Scale [10], we elected to retain the complete set of items. To facilitate meaningful interpretation of scores, all items were linked to the Toolbox metric, such that T scores ($M = 50$ and SD = 10) were comparable and representative of the US 2010 general population. This was accomplished by following the multi-method linking procedure described by PROsetta Stone investigators [41]. The resulting Stocking–Lord linking constants ($A = 1.094$ and $B = -0.507$) were applied to the PROMIS item parameters to place them on the Toolbox metric.

Identifying a short form and examining preliminary validity

Of particular relevance for identifying the "best" items for short forms was the information accounted for by each item across the general self-efficacy continuum. Based on information function and content considerations (capturing a conceptual range of general self-efficacy beliefs), we identified the "best" 4-item short form (Table 2) to go along with the standard 10-item bank. Two of the four items we selected

had S-X^2 values suggestive of possible poor fit. This might have been due to higher frequencies of endorsement relative to the average of the other items [39]. On balance, the strength of those items relative to information function and content validity merited inclusion on the short form.

Both the short form and the item bank demonstrated excellent internal consistency reliability, with coefficient αs = .88 and .94, respectively. Table 3 presents the bivariate correlations among the PROMIS General Self-Efficacy Short Form and Item Bank with related constructs (optimism, success expectancies), the parallel Toolbox measure, and the PROMIS Global-10. All correlations with the PROMIS Self-Efficacy measures were significant ($p < .001$) with $r \geq 0.39$. Not surprisingly, the highest correlations were found between the PROMIS General Self-Efficacy Short Form and Item Bank to the Toolbox Self-Efficacy Item Bank, $r = .85$ and .87, respectively (Table 3).

Conclusions

The PROMIS General Self-Efficacy measure demonstrated sufficient unidimensionality and displayed good internal consistency reliability, model fit, and convergent validity. This is the first report summarizing the psychometric properties of this addition to the PROMIS "family" of measures. In comparison to existing measures of general self-efficacy [10, 42, 43], these data describe the first study that has leveraged IRT with a large, diverse general population sample to refine assessment of this important, patient-centered construct for understanding healthy adaptation to acute and chronic illness.

The PROMIS General Self-Efficacy Item Bank performed equal to or slightly better than the NIH Toolbox Self-Efficacy Item Bank on all but one quantitative index. Specifically, PROMIS demonstrated less skew, had higher item-total correlations, accounted for more explained variance

Table 2 PROMIS® general self-efficacy

Item contents		Mean GRM threshold	X^2	d.f.	Probability
1[a]	I can manage to solve difficult problems if I try hard enough	− 1.2986	107.45	65	0.0007
2	If someone opposes me, I can find the means and ways to get what I want	− 0.7184	80.72	87	0.6696
3	It is easy for me to stick to my aims and accomplish my goals	− 0.9163	82.08	90	0.7121
4[a]	I am confident that I could deal efficiently with unexpected events	− 1.0403	81.34	72	0.2109
5	Thanks to my talents and skills, I know how to handle unexpected situations	− 1.0546	89.75	74	0.1025
6	I can solve most problems if I try hard enough	− 1.2669	118.95	71	0.0003
7	I stay calm when facing difficulties because I can handle them	− 0.9625	124.47	82	0.0017
8	When I have a problem, I can find several ways to solve it	− 1.1288	59.64	72	0.8509
9[a]	If I am in trouble, I can think of a solution	− 1.1886	97.80	66	0.0067
10[a]	I can handle whatever comes my way	− 1.0166	74.63	70	0.3297

Item content is reprinted with permission of the PROMIS Health Organization and not for reproduction. All PROMIS materials are ©2008–2019 PROMIS Health Organization. PROMIS is a registered trademark of HHS

GRM graded response model

[a]Short form item

Table 3 Reliabilities and bivariate correlations

	Cronbach's α	Life Orientation Test-Revised (LOT-R)	Generalized Expectancy for Success Scale	NIH Toolbox® Self-Efficacy Item Bank	PROMIS® Global-10: Mental Health	PROMIS® Global-10: Physical Health
PROMIS® General Self-Efficacy Short Form	0.88	0.56	0.52	0.85	0.53	0.39
PROMIS® General Self-Efficacy Item Bank	0.94	0.58	0.55	0.87	0.56	0.39

All Pearson correlations are significant at the $p < .001$ level

in EFA, and had comparable fit statistics in CFA compared to NIH Toolbox Self-Efficacy items with the exception of a slightly poorer fit for the RMSEA. Despite this, our data suggested that both measures are sufficiently unidimensional and locally independent, essential characteristics for good measurement within an IRT framework [23]. Given the performance of the NIH PROMIS General Self-Efficacy items, the alignment of its confidence response options with self-efficacy theory [21] and patient preference [22], and the match of the response options with existing PROMIS context-specific, self-efficacy measures [19], the PROMIS General Self-Efficacy Item Bank and Short Form provide an important self-efficacy assessment option.

The PROMIS® General Self-Efficacy items were successfully calibrated along the same metric and were free of DIF. By calibrating these items along the same metric, the item bank can be used as a CAT, minimizing respondent burden without sacrificing measurement precision. All items discriminated quite well, suggesting that they can accurately

assess differences between individuals who vary in general self-efficacy across the range of the construct. In addition, since there was no evidence of item bias (i.e., DIF), the items appear to function well for diverse groups of people with respect to age, gender, education, and race. Further, these items are calibrated along the same metric as the NIH Toolbox Self-Efficacy Item Bank, meaning that the scores are linked to a robust norming sample that is representative of the US 2010 Census [44]. As with all PROMIS measures, the PROMIS General Self-Efficacy measures use a *T* score with a mean of 50 and a standard deviation of 10, with higher scores indicating more of the underlying construct. This metric facilitates easy and understandable interpretation of scores.

The PROMIS General Self-Efficacy Short Form and PROMIS General Self-Efficacy Item Bank also demonstrated excellent psychometric properties when evaluated using classical test theory approaches. Specifically, both measures were highly reliable as evidenced by their excellent

internal consistency scores. Similarly, convergent validity correlations with related constructs such as optimism and positive expectancies were large and in the expected direction [45, 46], and the PROMIS General Self-Efficacy measures were highly correlated with the Toolbox version of the same construct [20], suggesting comparable approaches to assessing this important construct. Lastly, bivariate correlations with the PROMIS Global-10 revealed expected large associations with mental health and moderate associations with physical health. This finding perhaps underscores the positive associations between general self-efficacy as an adaptive personality trait and its connection with emotional well-being and, to a lesser extent, physical well-being [47–51].

Some limitations of this study are worth noting. First, although the measures are designed to be used in healthy individuals and those with a range of acute and chronic illnesses in a longitudinal fashion, this initial calibration and validation approach focused on a large, cross-sectional, general population sample. Future testing using a longitudinal design can assess the stability of general self-efficacy as assessed by this measure. Second, having a wider range of validation measures for evaluating convergent and discriminant validity would be beneficial. We intentionally kept our measurement battery to a modest length to minimize the potential for respondent fatigue that might compromise the validity of participants' responses. Future testing that compares the PROMIS General Self-Efficacy measures to other indices of related constructs such as mastery and control [52], autonomy [2], psychological hardiness [53], and resilience [54] measures would be informative.

Additional future directions for this work are to expand validity testing, examine complementary assessment strategies with domain-specific self-efficacy, identify cut-off thresholds for important differences, and consider the added value of standard setting applications with this measure to facilitate the clinical utility of scores [55]. For validity testing, we need to explore how this measure performs in clinical settings and its added value of assessing the PROMIS General Self-Efficacy Item Bank alongside PROMIS Self-Efficacy for Managing Chronic Conditions Item Banks. As more of a trait-based factor, PROMIS General Self-Efficacy may function as an important moderator of healthy adaptation to illness. For important differences and related tasks of standard setting, we hope to identify the minimally important difference on the PROMIS General Self-Efficacy measures that correlate with clinically significant outcomes, and to identify optimal levels of general self-efficacy for healthy functioning and/or mastery.

In summary, the PROMIS General Self-Efficacy Item Bank and Short Form are psychometrically sound measures of global self-efficacy. They provide robust assessments of an important, patient-centered construct with significant health relevance. These measures improve upon the existing measurement landscape of general self-efficacy through their integration and application of IRT, and they provide an important complement to existing PROMIS measures of self-efficacy for managing chronic conditions. Further psychometric testing will help evaluate the utility of this measurement tool in patients with chronic health conditions.

Compliance with ethical standards

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

Ethical approval All procedures performed involving human participants were in accordance with the Ethical Standards of the Institutional Review Board of Northwestern University and with the 1964 Helsinki Declaration and its later amendments.

Informed consent Informed consent was obtained from all participants in the study.

References

1. Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.
2. Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.
3. Carver, C. S., & Scheier, M. (1981). *Attention and self-regulation*. New York: Springer.
4. Rahman, A., et al. (2008). Factors affecting self-efficacy and pain intensity in patients with chronic musculoskeletal pain seen in a specialist rheumatology pain clinic. *Rheumatology, 47*(12), 1803–1808.
5. Schwerdtfeger, A., Konermann, L., & Schönhofen, K. (2008). Self-efficacy as a health-protective resource in teachers? A biopsychological approach. *Health Psychology, 27*(3), 358–368.
6. Jones, F., Partridge, C., & Reid, F. (2008). The Stroke Self-Efficacy Questionnaire: Measuring individual confidence in functional performance after stroke. *Journal of Clinical Nursing, 17*(7b), 244–252.
7. Merluzzi, T. V., et al. (2001). Self-efficacy for coping with cancer: Revision of the Cancer Behavior Inventory (version 2.0). *Psychooncology, 10*(3), 206–217.
8. Merluzzi, T. V., et al. (2018). Self-efficacy for coping with cancer: Revision of the Cancer Behavior Inventory (Version 3.0). *Psychol Assessment, 30*(4), 486–499.
9. Sheeran, P., et al. (2016). The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: A meta-analysis. *Health Psychology, 35*(11), 1178–1188.
10. Schwarzer, R., & Jerusalem, M. (1995). Generalized self efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology* (pp. 35–37). Windsor: NFER-Nelson.

11. Schwarzer, R., & Warner, L. M. (2013) Perceived self-efficacy and its relationship to resilience. In *Resilience in children, adolescents, and adults* (pp. 139–150). New York: Springer.

12. Courneya, K. S., et al. (2012). Effects of supervised exercise on motivational outcomes and longer-term behavior. *Medicine and Science in Sports and Exercise, 44*(3), 542–549.

13. Demark-Wahnefried, W., et al. (2003). Design of FRESH START: A randomized trial of exercise and diet among cancer survivors. *Medicine and Science in Sports and Exercise, 35*(3), 415–424.

14. Chesney, M. A., et al. (2006). A validity and reliability study of the coping self-efficacy scale. *British Journal of Health Psychology, 11*(Pt 3), 421–437.

15. Lorig, K. R., et al. (2001). Effect of a self-management program on patients with chronic disease. *Effective Clinical Practice, 4*(6), 256–262.

16. Cella, D., et al. (2010). Initial adult health item banks and first wave testing of the Patient-Reported Outcomes Measurement Information System (PROMIS™) Network: 2005–2008. *Journal of Clinical Epidemiology, 63*(11), 1179.

17. Cella, D., et al. (2016). PROMIS fatigue item bank had clinical validity across diverse chronic conditions. *Journal of Clinical Epidemiology, 73,* 128–134.

18. Cook, K. F., et al. (2016). PROMIS measures of pain, fatigue, negative affect, physical function, and social function demonstrated clinical validity across a range of chronic conditions. *Journal of Clinical Epidemiology, 73,* 89–102.

19. Gruber-Baldini, A. L., et al. (2017). Validation of the PROMIS® measures of self-efficacy for managing chronic conditions. *Quality of Life Research*. https://doi.org/10.1007/s11136-017-1527-3.

20. Kupst, M. J., et al. (2015). Assessment of stress and self-efficacy for the NIH Toolbox for Neurological and Behavioral Function. *Anxiety Stress and Coping, 28*(5), 531–544.

21. Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.

22. Salsman, J. M., et al. (2018). Refining and supplementing candidate measures of psychological well-being for the NIH PROMIS®: Qualitative results from a mixed cancer sample. *Quality of Life Research*. https://doi.org/10.1007/s11136-018-1896-2.

23. PROMIS Health Organization and PROMIS Cooperative Group. (2013). *PROMIS® instrument development and validation: Scientific standards version 2.0 (revised May 2013)*. Retrieved September 19, 2013, from http://www.nihpromis.org/Documents/PROMISStandards_Vers2.0_Final.pdf.

24. Liu, H., et al. (2010). Representativeness of the PROMIS Internet Panel. *Journal of Clinical Epidemiology, 63*(11), 1169–1178.

25. Roster, C. A., et al. (2004). A comparison of response characteristics from web and telephone surveys. *International Journal of Market Research, 46*(3), 359–374.

26. Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*(6), 1063–1078.

27. Fung, H. H., & Carstensen, L. L. (2004). Motivational changes in response to blocked goals and foreshortened time: Testing alternatives to socioemotional selectivity theory. *Psychology and Aging, 19*(1), 68–78.

28. Hale, W. D., Fiedler, L. R., & Cochran, C. D. (1992). The revised Generalized Expectancy for Success Scale: A validity and reliability study. *Journal of Clinical Psychology, 48*(4), 517–521.

29. Hays, R. D., et al. (2009). Development of physical and mental health summary scores from the Patient Reported Outcomes Measurement Information System (PROMIS) global items. *Quality of Life Research, 18*(7), 873–880.

30. Pilkonis, P. A., et al. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS): Depression, anxiety, and anger. *Assessment, 18*(3), 263–283.

31. Reeve, B. B., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care, 45*(5 Suppl 1), S22–S31.

32. R Core Development Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.r-project.org/.

33. Revelle, W. (2018). *Psych: Procedures for personality and psychological research. (R Package Version 1.8.12)*. (Computer Software). Evanston, IL: Northwestern University. https://cran.r-project.org/web/packages/psych/index.html.

34. Schmid, J., & Leiman, J. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53–61.

35. Reise, S. P., et al. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5–26.

36. Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores.* Psychometrika monograph supplement, no. 17. Richmond, VA: Psychometric Society.

37. Li, C., Thissen, D., & du Toit, S. (2011). *IRTPRO 2.1 for Windows*. Lincolnwood, IL: Scientific Software International, Inc.

38. Orlando, M., & Thissen, D. (2003). Further examination of the performance of S-X^2, an item fit index for dichotomous item response theory models. *Applied Psychological Measurement, 27,* 289–298.

39. Kang, T., & Chen, T. T. (2011). Performance of the generalized S-X^2 item fit index for the graded response model. *Asia–Pacific Education Review, 12*(1), 89–96.

40. Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). LORDIF: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*(8), 1–30.

41. Choi, S. W., et al. (2014). Establishing a common metric for depressive symptoms: Linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychological Assessment, 26*(2), 513–527.

42. Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods, 4*(1), 62–83.

43. Sherer, M., et al. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports, 51*(2), 663–671.

44. Babakhanyan, I., et al. (2018). National Institutes of Health Toolbox Emotion Battery for English- and Spanish-speaking adults: Normative data and factor-based summary scores. *Patient Related Outcome Measures, 9,* 115–127.

45. Kostka, T., & Jachimowicz, V. (2010). Relationship of quality of life to dispositional optimism, health locus of control and self-efficacy in older subjects living in different environments. *Quality of Life Research, 19*(3), 351–361.

46. Magaletta, P. R., & Oliver, J. M. (1999). The hope construct, will, and ways: Their relations with self-efficacy, optimism, and general well-being. *Journal of Clinical Psychology, 55*(5), 539–551.

47. Bonsaksen, T., et al. (2018). Self-diagnosed depression in the Norwegian general population—Associations with neuroticism, extraversion, optimism, and general self-efficacy. *BMC Public Health, 18*(1), 1076.

48. Hajek, A., & Hans Helmut, K. (2017). The role of self-efficacy, self-esteem and optimism for using routine health check-ups in a population-based sample. A longitudinal perspective. *Preventive Medicine, 105,* 47–51.

49. Schwarzer, R., et al. (1997). The assessment of optimistic self-beliefs: Comparison of the German, Spanish, and Chinese versions of the general self-efficacy scale. *Applied Psychology, 46*(1), 69–88.

50. Scholz, U., et al. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*(3), 242.

51. Roddenberry, A., & Renk, K. (2010). Locus of control and self-efficacy: Potential mediators of stress, illness, and utilization of health services in college students. *Child Psychiatry and Human Development, 41*(4), 353–370.

52. Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*(4), 719–727.

53. Kobasa, S. C. (1982). The hardy personality: Toward a social psychology of stress and health. *Social Psychology of Health and Illness, 4,* 3–32.

54. Connor, K. M., & Davidson, J. R. (2003). Development of a new resilience scale: The Connor–Davidson resilience scale (CD-RISC). *Depression and Anxiety, 18*(2), 76–82.

55. Cella, D., et al. (2014). Setting standards for severity of common symptoms in oncology using the PROMIS item banks and expert judgment. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 23*(10), 2651–2661.