ORIGINAL ARTICLE

University of Washington Self-Efficacy Scale: A New Self-Efficacy Scale for People With Disabilities

Dagmar Amtmann, PhD, Alyssa M. Bamer, MPH, Karon F. Cook, PhD, Robert L. Askew, MPH, Vanessa K. Noonan, PhD, Jo Ann Brockway, PhD

ABSTRACT. Amtmann D, Bamer AM, Cook KF, Askew RL, Noonan VK, Brockway JA. University of Washington Self-Efficacy Scale: a new self-efficacy scale for people with disabilities. Arch Phys Med Rehabil 2012;93:1757-65.

Objective: To develop a self-efficacy scale for people living with multiple sclerosis (MS) and spinal cord injury (SCI) that can be used across diagnostic conditions.

Design: The scale was developed using modern psychometric methods including item response theory. Items were administered at 3 time-points of a longitudinal survey of individuals with MS and SCI.

Setting: Survey participants with MS were recruited from the National MS Society, and participants with SCI were recruited from the Northwest Regional Spinal Cord Injury Model System and the Shepherd Center at the Virginia Crawford Research Institute in Atlanta, GA.

Participants: Adults aged 18 years and older reporting a definitive diagnosis of MS (N=473) or SCI (N=253).

Interventions: None.

Main Outcome Measures: Evaluation of the new self-efficacy measure called the University of Washington Self-Efficacy Scale (UW-SES) included comparisons with the Chronic Disease Self-Efficacy Scale and other patient-reported outcome measures.

Results: UW-SES has excellent psychometric properties including well-functioning response categories, no floor effects, and low ceiling effects. A long form (17 items) and a short form (6 items) are available. The correlation between the score on the newly developed scale and the Chronic Disease Self-Efficacy Scale was high (.83), providing support for convergent validity. Higher self-efficacy scores were statistically significantly associated with better mental health, better physical health, less fatigue, less stress, less pain interference, less pain, fewer sleep problems, and lower depressive symptoms.

Conclusions: The UW-SES is a psychometrically sound instrument for measuring self-efficacy, validated in MS and SCI, and can be used across both conditions. Both the long form and the short form are available free of charge.

Key Words: Multiple sclerosis; Patient outcome assessment; Rehabilitation; Self-efficacy; Spinal cord injuries.

SELF-EFFICACY IS THE belief in one's ability to produce the effects or outcomes one wants.[1] The construct of self-efficacy is a core component of social cognitive theory, in which psychosocial functioning is determined by reciprocal interactions between an individual's personal (cognitive, biological, and affective) factors, his/her behavior, and the environment in which he/she functions.[1] People develop perceptions about their capabilities, and these perceptions mediate future behavior. Self-efficacy beliefs impact behavior through several avenues. Self-efficacy beliefs influence the course of action an individual chooses; that is, most people choose a course of action in which they feel competent rather than one in which they do not. In addition, one's belief in one's ability to succeed influences the amount of effort expended, the extent of stress experienced, and the degree of perseverance in the face of difficulties.

There is strong theoretical support for the relationship between self-efficacy and psychological well-being, and their association has great practical importance. This association has been documented for self-efficacy in numerous chronic conditions.[2,3] For instance, a behavior-specific measure of perceived self-efficacy in arthritis was found to be highly correlated with functioning as measured by the Stanford Health Assessment Questionaire and strongly inversely related to depression.[4] Self-efficacy has been shown to be a strong predictor of health behaviors,[5] and as such it can be an important modulator of the experience of chronic illness. Studies have shown that specific self-efficacy is closely related to important outcome measures such as subjective well-being,[6] functional recovery after hip fracture,[7] pain,[8] employment outcome from psychiatric rehabilitation,[9] coping in patients who had tumor surgery,[10] and psychological well-being after spinal

List of Abbreviations

CFA	confirmatory factor analysis
DIF	differential item functioning
IRT	item response theory
MS	multiple sclerosis
PROMIS	Patient Reported Outcomes Measurement Information System
SCI	spinal cord injury
SF-8	Medical Outcomes Study Short-Form 8-Item Health Survey
TI	Test-Information
UW-SES	University of Washington Self-Efficacy Scale

From the Department of Rehabilitation Medicine, University of Washington, Seattle, WA (Amtmann, Bamer, Askew, Brockway); Northwestern University Feinberg School of Medicine, Chicago, IL (Cook); and Department of Orthopaedics, University of British Columbia, BC, Canada (Noonan).

Presented in part to the Consortium of Multiple Sclerosis Centers, June 4, 2010, San Antonio, TX, and June 3, 2011, Montreal, Canada.

Supported by the Department of Education, National Institute on Disability and Rehabilitation Research (grant nos. H133B031129 and H133B080025), and the National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institute of Health (grant no. 5U01AR052171). However, these contents do not necessarily represent the policy of the Department of Education, and one should not assume endorsement by the federal government.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit on the authors or on any organization with which the authors are associated.

Reprint requests to Dagmar Amtmann, PhD, Department of Rehabilitation Medicine, University of Washington, Box 356490, Seattle, WA 98195, e-mail: *dagmara@u.washington.edu*.

In-press corrected proof published online on Jun 22, 2012, at www.archives-pmr.org.

0003-9993/12/9310-00154$36.00/0
http://dx.doi.org/10.1016/j.apmr.2012.05.001

cord injury (SCI).[11]Interventions have also been developed to increase self-efficacy with the goal of improving chronic disease outcomes.[12]

Numerous self-efficacy scales have been constructed to measure either generalized self-efficacy or self-efficacy specific to a content area. For individuals with multiple sclerosis (MS) or SCI, common challenges are often based on limitations in mobility, changes in appearance, decreased sensation, changes in bowel and bladder function, chronic pain, and ongoing medical complications. These challenges may affect an individual's confidence in his/her ability to achieve desired goals such as having meaningful relationships, a sense of well-being, or the ability to manage health issues. The purpose of this study was to develop a self-efficacy instrument to measure self-perceived confidence in managing challenges related to MS and SCI that could be used across several disability groups. In addition, we wanted to develop an instrument using modern psychometric approaches, including the item response theory (IRT), to take advantage of the applications of IRT, such as computerized adaptive testing, targeted short forms, and ability to examine differential item functioning (DIF). IRT also facilitates the examination of psychometric functioning of the items and the scale in a way that is not available within the classical test theory framework. In particular, in an IRT framework, reliability values are estimated for every level of the trait being measured, while in classical test theory, reliability is assumed to be the same for an entire scale. However, scales typically measure different levels of a trait with different precision; that is, the precision is usually higher around the mean and lower at both ends of the continuum.

The overall goal of the study was to develop a scale that (1) measures self-efficacy specific to chronic conditions, such as MS and SCI, (2) could be used across different disability groups, (3) could be applicable to other disabilities or conditions, and (4) is scored and examined using IRT methodology.

METHODS

Development of Initial Item Pool

A literature search was conducted to identify existing measures of self-efficacy suitable for use in populations of individuals with disabilities. Several instruments were reviewed with respect to the construct measured, factor structure, reliability, and validity. Most instruments measured self-efficacy specific to chronic diseases such as MS,[13] SCI,[14,15] MS or SCI,[16] epilepsy,[2] arthritis,[4] and chronic disease in general.[17] One instrument measured general self-efficacy.[18]

Initial items were developed to assess respondents' confidence in their ability to meet commonly reported challenges in MS and SCI. Several items were adapted from the Chronic Disease Self-Efficacy Scale (eg, How confident are you that you can keep the physical discomfort or pain of your disease from interfering with the things you want to do?).[19] Additional new items were developed with input from health care professionals with extensive experience in treating people with MS and SCI. Candidate items were administered repeatedly to participants with MS and SCI in a longitudinal study. Items were modified as analyses indicated after each administration. However, individuals with MS and SCI were not specifically interviewed about item or scale wording or content coverage.

Research Participants

Study participants were involved in an ongoing longitudinal study of self-reported symptoms and quality-of-life indicators in MS and SCI. Previous publications have described recruitment procedures for both the MS[20-23] and SCI[24] participants in detail. Briefly, participants with MS were recruited from the mailing list of the Western Washington chapter of the National MS Society, and participants with SCI were recruited from the Northwest Regional Spinal Cord Injury Model System (http://sci.washington.edu/) and the Shepherd Center at the Virginia Crawford Research Institute in Atlanta, Georgia. Study participation was limited to eligible adults aged 18 years and older reporting a definitive diagnosis of MS or SCI. Data for this study were collected in 2006 and 2007. A total of 1891 (MS=1271, SCI=620) individuals completed the first survey, and a subset of 926 (MS=562, SCI=364) was randomly selected to continue in the longitudinal survey study, which involved completion of surveys every 4 months. The Human Subjects Division at the University of Washington approved all study procedures, and participants were paid $25 for completing the surveys at each time-point.

Research Measures

All measures, including candidate items, were administered via self-report survey, and domains assessed in the survey primarily focused on the common symptoms of MS and SCI, including pain, fatigue, depression, anxiety, mobility changes, and sleep problems. Information on age, sex, and other demographic characteristics was also collected. Each of the following scales were scored according to published recommendations: the Chronic Disease Self-efficacy Scale,[19] the Medical Outcomes Study Short-Form 8-Item Health Survey (SF-8),[25] the Modified Fatigue Impact Scale,[26] the Perceived Stress Scale,[27,28] the Medical Outcomes Study Sleep scale,[29] the Pain Impact Questionnaire-6,[30] the Patient Health Questionnaire 9-item depression scale,[31,32] the Patient Reported Outcomes Measurement Information System (PROMIS) anxiety short form,[33] the PROMIS satisfaction with social roles short form,[33] the PROMIS satisfaction with discretionary activities short form,[33] and a single item measuring average pain intensity on a scale from 1 to 10.[34] Scores on the SF-8 mental and physical subscales range from 0 to 100, with higher scores indicating better health. The Modified Fatigue Impact Scale total scores range from 0 to 84, with higher scores indicating higher fatigue. The Perceived Stress Scale 4-item short-form scores range from 0 to 16, with higher scores indicating more stress. Scores from the Medical Outcomes Study Sleep scale sleep problems index 9 subscale range from 0 to 100, with higher scores indicating more sleep problems. The Pain Impact Questionnaire-6 scores range from 40 to 78, with higher scores indicating greater pain interference. The Patient Health Questionnaire 9-item depression scale total score was used to measure depression, and scores range from 0 to 27, with higher scores indicating more depressive symptoms. Scores on the 3 PROMIS measures range from 0 to 100, with the U.S. general population mean centered at 50. Measures collected for validity assessment were from the fourth survey time-point (12mo) or the sixth survey time-point (20mo) with the exception of the PROMIS social scales, which were from the fifth survey time-point (16mo).

Candidate Item Pool

Candidate self-efficacy items were administered during the first, second (4mo later), and fourth time-points (12mo later), with item modification following the first 2 administrations. The first survey contained 14 candidate items, the second contained 15 candidate items, and the fourth survey (third administration) contained 20 candidate items. Final item selection and scale scoring were carried out using data from the fourth time-point. Preliminary analyses of the 20 candidate

items identified 3 items with problematic responses (ie, significant missing data/nonresponse or dichotomous response patterns). Because the authors felt that 2 of these items covered important content areas (ie, sexual and intimate relationships), and although they were excluded from analyses and scoring, they were retained in the scale. These 2 items are not intended to be scored but can be used for clinical and descriptive purposes by clinicians and researchers. One additional item was dropped because of local dependency identified during the examination of the factor structure. All subsequent analyses were completed using the remaining 17 candidate items.

Factor Structure

During scale development, repeated examination of the factor structure is critical to ensure that a single latent construct drives the variance in item responses. Confirmatory factor analysis (CFA) was carried out on the 17 candidate items using Mplus software (V5.21).[a] One-factor CFA was applied, and fit statistic values were compared with published criteria set forth by Bentler,[35] Hu and Bentler,[36,37] McDonald,[38] and others.[39-41] The criteria used were the comparative fit index of more than .95, root mean square error of approximation of less than .06, and the Tucker Lewis index of more than .95. Because CFA fit values have been found to be sensitive to data distribution and the number of items,[42] when the criteria for acceptable fit with a 1-factor CFA are not met, a bifactor model[43] can be used to evaluate the degree of unidimensionality. In the bifactor model, all items load on a single, general factor and subsets of items load on subdimensions (group factors). Modeling the data using the bifactor model generates an estimate of the amount of variance accounted for by the group factors compared with the variance accounted for by the general (self-efficacy) factor.[44] In this study, the items were assigned to the 2 group factors using results from an exploratory factor analysis.

IRT Calibration and Fit

Responses to the 17 candidate items were modeled using Samejima's[45] 2 parameter polytomous graded response model with Multilog, Version 7.03.[b] This IRT model allowed for the comparison of the patterns of actual participant responses with those predicted by the model. Fit to the graded response model was then evaluated using the SAS macro IRTFIT,[c] and items would be considered to exhibit poor fit if S-χ^2 (Pearson χ^2) and S-G^2 (likelihood ratio G^2 statistic)[46,47] were $\leq$.01.

Scoring

Scoring tables for the final University of Washington Self-Efficacy Scale (UW-SES) and short form were developed using item parameters generated from the graded response model. IRT scores are on the theta metric, and as such, scale scores can take on negative values that are not intuitive to users. To make scale scores more user-friendly, theta scores were then transformed to a T-scale score with a mean of 50 and an SD of 10. The program IRT score[d] was used to generate a summary score to theta conversion table using the item parameters from Multilog. Subsequent reliability analyses were carried out using the T scores.

Reliability

In IRT, reliability is evaluated by Test-Information (TI) that is computed for each level of the trait (ie, self-efficacy). The standard error of measurement is inversely related to TI. In

Table 1: Demographic Characteristics of a Community Sample of Individuals With MS (n=473) or SCI (n=253)

Variable	MS	SCI
Age (y)	52.3±10.9 [21–82]	47.1±14.3 [19–85]
Duration of disease (y)	14.5±9.9 [2–61]	13.4±10.2 [2–57]
Sex		
Women	391 (82.7)	94 (37.2)
Men	82 (17.3)	159 (62.9)
Race*		
Caucasian	458 (96.8)	211 (83.7)
Native American or Alaska Native	16 (3.4)	9 (3.6)
Asian	4 (0.9)	6 (2.4)
African-American	9 (1.9)	29 (11.5)
Education completed		
<High school	4 (0.8)	13 (5.1)
High school	64 (13.5)	50 (19.8)
Vocational/some college	180 (38.1)	93 (36.8)
Bachelor degree	140 (29.6)	68 (26.9)
Professional/graduate	85 (18.0)	29 (11.5)
Employment status		
Employed 20+ h/wk	155 (32.8)	71 (28.1)
Employed <20 h/wk	21 (4.4)	11 (4.4)
Retired	70 (23.5)	30 (17.8)
Marital status		
Married/live with significant other	323 (68.3)	118 (46.6)
Separated/divorced	71 (15.0)	58 (22.9)
Never married	25 (5.3)	20 (7.9)
Widowed	15 (3.2)	5 (2.0)
SF-8 Mental score	46.6±10.2	48.6±10.4
SF-8 Physical score	40.3±10.2	38.9±9.6

NOTE. Values are n (%), mean ± SD, and [range].
*Individuals were allowed to choose multiple answers.

addition to the classical test theory reliability measure (Chronbach's α), TI and standard error of measurement were examined to evaluate the UW-SES's reliability and precision.

Validity Assessment

The construct validity of the UW-SES and the short form was assessed by comparing scale scores with those of multiple related measures collected at the 12-month and 20-month surveys. Pearson's r correlation coefficients were calculated between the UW-SES final T scores and the Stanford self-efficacy summary scores. A linear regression adjusted for the effects of age and sex was performed to assess convergent and divergent validity with measures that may theoretically be influenced by an individual's disease self-efficacy. Measures included in the cross-sectional regression analyses were the SF-8 mental and physical subscales, total fatigue, perceived stress, sleep prob-

Table 2: List of Self-Efficacy Scale Item Stems With Their Corresponding Parameter Estimates and Model Fit Statistics

Item Stem	Parameter Estimates					S-χ^2		S-G^2	
How Confident Are You That	Slope	Threshold 1	Threshold 2	Threshold 3	Threshold 4	χ^2	χ^2 *P*	G^2	G^2 *P*
1. You can keep the fatigue caused by your MS/SCI from interfering with the things you want to do?	2.47	−1.75	−0.73	0.17	1.21	56.78	0.85	61.85	0.72
2. You can keep the physical discomfort of your MS/SCI from interfering with the things you want to do?*	2.58	−1.92	−0.75	0.05	1.26	53.16	0.93	55.39	0.90
3. You can keep the pain of your MS/SCI from interfering with the things you want to do?	2.30	−1.92	−0.86	−0.01	1.06	63.94	0.68	65.68	0.62
4. You can keep the emotional distress caused by your MS/SCI from interfering with the things you want to do?	2.49	−2.35	−1.13	−0.30	0.73	61.82	0.69	64.37	0.60
5. You can keep any other symptoms or health problems you have from interfering with the things you want to do?	2.12	−2.44	−1.00	−0.01	1.29	50.16	0.97	51.23	0.96
6. You can do things other than just taking medication to reduce how much your MS/SCI affects your everyday life?	1.83	−2.13	−0.92	−0.10	1.09	69.15	0.61	82.45	0.21
7. You can keep your MS/SCI from interfering with managing your affairs?	3.40	−1.93	−1.04	−0.21	0.70	67.93	0.51	68.34	0.50
8. You can keep your MS/SCI from interfering with family relationships?	3.08	−2.27	−1.20	−0.42	0.43	52.65	0.88	54.71	0.84
9. You can keep your MS/SCI from interfering with close friendships?	3.21	−2.24	−1.13	−0.36	0.46	31.96	1.00	33.80	1.00
10. You can keep your MS/SCI from interfering with your ability to deal with unexpected events?*	3.08	−1.90	−0.84	0.05	1.15	52.53	0.87	52.67	0.86
11. You can keep your MS/SCI from interfering with your ability to interact socially?*	3.83	−2.03	−0.97	−0.10	0.78	34.87	1.00	35.20	1.00
12. You can keep your MS/SCI from being the center of your life?*	2.56	−1.95	−1.08	−0.25	0.76	38.22	1.00	39.20	1.00
13. You can keep your MS/SCI from interfering with having a fulfilling life?	2.97	−1.83	−1.01	−0.26	0.71	46.61	0.97	48.01	0.96
14. You can, using all the resources available to you, minimize the occurrence of MS-/SCI-related complications (such as bladder accidents or falls)?	1.78	−2.56	−1.36	−0.32	1.03	90.34	0.08	98.54	0.02
15. You can bounce back from frustration, discouragement, or disappointment that MS/SCI may cause you?*	2.42	−2.73	−1.43	−0.38	0.81	57.95	0.72	59.62	0.67
16. You can, using all the resources available to you, successfully manage your medication needs?	1.57	−3.25	−2.27	−1.38	−0.48	74.49	0.13	76.04	0.11
17. You can figure out effective solutions to MS-/SCI-related issues that come up?*	2.31	−3.03	−1.57	−0.60	0.61	81.35	0.10	84.28	0.06
18. You can keep your MS/SCI from interfering with having an emotionally intimate relationship with a spouse or partner?†				NA					
19. You can keep your MS/SCI from interfering with having a satisfying sexual relationship?†				NA					

Abbreviation: NA, not applicable.
*Item is included in the 6-item short form.
†Item included for clinical or descriptive purposes only and is not included in the overall scale score.

lems, average pain intensity, pain interference, and depression. To examine how self-efficacy predicted health at the 20-month follow-up, analyses included measures of overall health, total fatigue, perceived stress, pain interference, and depression. In addition to the regression analyses, Pearson's r correlation coefficients were calculated between the scores on the UW-SES and related constructs to compare the associations found in this study with previously reported results. We also compared the mean self-efficacy levels by marital and employment status with previously published results.

Differential Item Functioning

To examine whether items performed similarly across sex, age ($<$58y vs $\geq$58y), and diagnostic subgroups (MS vs SCI), a DIF analysis was performed. The cutoff of 58 years was selected because it was the highest age for which at least 200 participants at that age or older were available.[48] Clinically meaningful DIF by diagnostic subgroups would suggest that some or all items may not be appropriate for both conditions (MS/SCI). The absence of meaningful DIF would provide support for its applicability for both populations. DIF was assessed using 3 statistical criteria: χ^2, changes in pseudo-R^2, and percentage change in regression coefficients. As χ^2 is known to be sensitive to sample size,[49] the alternative fit measures have been used to distinguish negligible from meaningful DIF. Following previously published guidelines, a change of less than .013 in pseudo-R^2 was considered negligible DIF,[50] as was a change of less than 5% to 10% in regression coefficients when group effects were added to the model.[51,52] DIF was assessed for each item using the LORDIF software package.[e]

Short Form

Authors examined each item with respect to the psychometric properties (eg, item difficulty and discrimination) and content coverage to select by consensus a subset of items for the short form. Short forms with different numbers of items were constructed to select the length of the instrument that balances brevity and precision. Because the final short form chosen for publication included 6 items, subsequent analyses used the 6-item short form.

RESULTS

Research Participants

A total of 1891 (MS=1271, SCI=620) individuals completed the first survey, 831 (MS=513, SCI=318) completed

Table 3: Conversion Table to Transform Self-Efficacy Scale Summary Scores to θ or T scores

Full 17-Item Scale Summary Score to T-Score Concordance Table						6-Item Short-Form Summary Score to T-Score Concordance Table		
Summary Score	θ Score	T Score	Summary Score	θ Score	T Score	Summary Score	θ Score	T Score
17	−3.46	15.40	52	−0.61	43.90	6	−3.00	20.0
18	−3.19	18.10	53	−0.56	44.40	7	−2.62	23.8
19	−3.00	20.00	54	−0.50	45.00	8	−2.35	26.5
20	−2.85	21.50	55	−0.45	45.50	9	−2.11	28.9
21	−2.71	22.90	56	−0.39	46.10	10	−1.90	31.0
22	−2.60	24.00	57	−0.34	46.60	11	−1.71	32.9
23	−2.49	25.10	58	−0.28	47.20	12	−1.53	34.7
24	−2.40	26.00	59	−0.22	47.80	13	−1.36	36.4
25	−2.31	26.90	60	−0.17	48.30	14	−1.20	38.0
26	−2.22	27.80	61	−0.11	48.90	15	−1.04	39.6
27	−2.14	28.60	62	−0.05	49.50	16	−0.89	41.1
28	−2.07	29.30	63	0.01	50.10	17	−0.74	42.6
29	−1.99	30.10	64	0.06	50.60	18	−0.59	44.1
30	−1.92	30.80	65	0.12	51.20	19	−0.44	45.6
31	−1.85	31.50	66	0.18	51.80	20	−0.29	47.1
32	−1.78	32.20	67	0.25	52.50	21	−0.14	48.6
33	−1.72	32.80	68	0.31	53.10	22	0.02	50.2
34	−1.65	33.50	69	0.37	53.70	23	0.18	51.8
35	−1.59	34.10	70	0.44	54.40	24	0.34	53.4
36	−1.53	34.70	71	0.50	55.00	25	0.51	55.1
37	−1.46	35.40	72	0.57	55.70	26	0.69	56.9
38	−1.40	36.00	73	0.64	56.40	27	0.90	59.0
39	−1.34	36.60	74	0.71	57.10	28	1.13	61.3
40	−1.28	37.20	75	0.79	57.90	29	1.41	64.1
41	−1.23	37.70	76	0.87	58.70	30	1.89	68.9
42	−1.17	38.30	77	0.95	59.50			
43	−1.11	38.90	78	1.04	60.40			
44	−1.05	39.50	79	1.13	61.30			
45	−1.00	40.00	80	1.24	62.40			
46	−0.94	40.60	81	1.36	63.60			
47	−0.88	41.20	82	1.50	65.00			
48	−0.83	41.70	83	1.66	66.60			
49	−0.77	42.30	84	1.89	68.90			
50	−0.72	42.80	85	2.26	72.60			
51	−0.66	43.40						

the second survey, 726 (MS=473, SCI=253) completed the fourth survey, and 671 individuals (MS=446, SCI=225) completed the sixth (20-mo) survey. A random subset of participants from the first survey was invited to continue in the remaining surveys; however, the decrease in participants between the second and sixth surveys was due to attrition. The attrition level was low (6% in MS and 11% in SCI over the 20-mo period), and the sample size at the last time-point was more than adequate for the analyses conducted.

The calibration sample was composed of the data collected at the fourth time-point. Individuals who completed the fourth survey time-point had a mean age of 50.5±12.4years, were 67% women (n=485), and had a mean disease duration of 14.1±10.0 years. Additional demographic information by the diagnostic condition can be found in table 1.

Factor Structure

The dimensionality of the 17-item bank was evaluated using CFA and bifactor analyses. Because the results using the 3 statistical criteria (comparative fit index, root mean square error of approximation, and Tucker Lewis index) were mixed, we fit the bifactor CFA that provided support for a sufficiently unidimensional factor structure. The general factor accounted for 59% of the total variance, and the group factors accounted for 13% of the total variance. The results suggested that the unidimensionality assumption of IRT was met.

IRT Calibration and Fit

Each of the 17 items fit the IRT-based model well at an alpha threshold of .01. Probability values for S-χ^2 and S-G^2 statistics ranged from .08 to 1.0 (mean=.71) and .03 to 1.0 (mean=.65), respectively. Item parameters and measures of fit are presented in table 2, along with each item stem.

Scoring

Scores on the transformed T-score metric for the 17-item UW-SES, which is centered on our sample, ranged from 24.0 to 72.6, with a mean of 50.0±9.7 (n=725). Scores on the 6-item short form ranged from 23.8 to 68.9, with a mean of 49.9±9.3 (n=726). Higher scores indicate higher self-efficacy. One person did not complete all the items and was excluded from the analyses. A summary score to theta and T-score conversion table is presented in table 3.

Reliability

Cronbach alpha for the 17-item UW-SES version was .96, and for the 6-item short form it was .90. Regarding TI, the 17-item scale provides substantial information across the range of self-efficacy observed in the sample as illustrated in figure 1. Three percent of the sample (n=24) endorsed "completely confident" for every item on the 17-item scale, while 6% of the sample (n=44) endorsed "completely confident" on the short form. This result suggests a small ceiling effect on both the full and short forms.

Validity

Correlation coefficients between the Chronic Disease Self-efficacy 6-item scale and the 17-item and 6-item forms of the UW-SES were high (.83 and .81, respectively), providing support for convergent validity. Table 4 presents the results of linear regression analyses examining the relationship between scores on the 17-item and 6-item forms with other measures hypothesized to be associated with self-efficacy. Both cross-sectionally and in predicting future scores, higher self-efficacy scores were found to be statistically significantly associated with better mental health, better physical health, less fatigue, less stress, less pain interference, and less depressive symptoms (all $P<.001$). In addition, cross-sectional analyses of sleep and average pain intensity indicate that higher self-efficacy is associated with fewer sleep problems and less pain intensity.

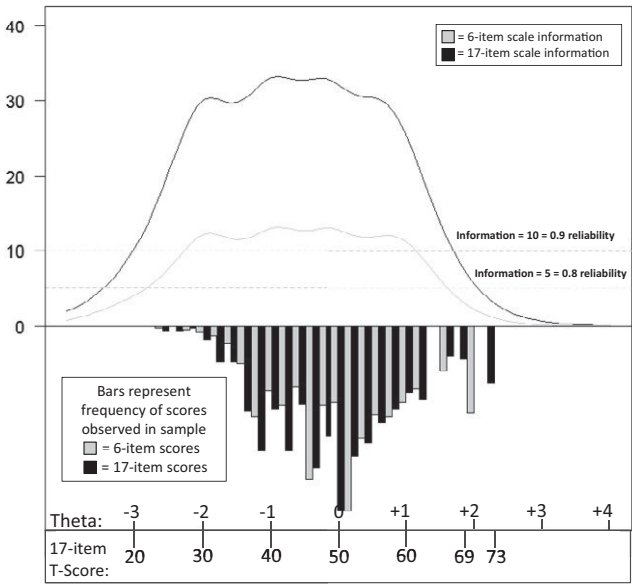

Fig 1. A 2-panel IRT-based information graph of the 17-item and 6-item self-efficacy scales with reliability levels of 0.8 and 0.9 marked. The panel above the line that represents 0 information is the IRT-based instrument information function that shows precision at different levels of self-efficacy. The score measures with a precision sufficient for individual comparisons (ie, above 0.9 reliability) from about −3 theta (T score of 20) to about 1.8 theta (T score of 68) for the long form and −2.3 theta (T score of 28) to about 1 theta (T score of 60) for the 6-item short form. This means that the score based on the 6-item short form for individuals who score at 1 SD above the mean (T score of 60) or higher is not precise enough to differentiate higher levels of self-efficacy at the individual level. The second panel below the 0 information line is a histogram of the calibration sample that shows the distribution of the scores in the sample that was used to calibrate the scale items to document that the sample included individuals across the whole continuum of self-efficacy.

The results of the correlation analyses between the UW-SES and scores on related domains found that the correlation between the 17-item UW-SES and depression was −.62 and for anxiety −.46. The correlation between the UW-SES and the physical subscale of SF-8 was positive and moderately large (r=.57). Those who were married or living with a significant other reported slightly higher mean self-efficacy score than did others (mean=51 vs 49). In addition, those who were employed reported a higher mean self-efficacy score than did those who were not (mean=48 vs 53). Age, duration of condition, and educational level had low correlations with the UW-SES score (−.04, .02, and .12). At the time-point after the UW-SES was administered, we measured social function and found a high correlation between the UW-SES score and the PROMIS satisfaction with social roles score (.68) and satisfaction with discretionary activities score (.70).

Differential Item Functioning

None of the 17 items had statistically significant DIF by disease status (MS, n=473; SCI n=253), age group (<58y, n=505; ≥58y, n=221), or by sex (man, n=241; woman, n=485) using any of the 3 recommended cutoff criteria.

Table 4: Linear Regression Results Between the 17- and 6-Item Short-Form T Scores and Scores on Other Outcome Measures Both Cross-Sectionally and in the Future*

	17-Item Scale						6-Item Short Form					
Measure	β	Standard Error	T Score	*P*	Adjusted Model R^2	N	β	Standard Error	T Score	*P*	Adjusted Model R^2	N
Cross-sectional												
SF-8 Mental	0.61	0.03	18.9	<.001	0.35	723	0.61	0.03	18.0	<.001	0.32	724
SF-8 Physical	0.59	0.03	19.2	<.001	0.38	723	0.59	0.03	18.4	<.001	0.36	724
Total fatigue	−1.28	0.05	−25.9	<.001	0.50	724	−1.30	0.05	−24.7	<.001	0.48	725
Perceived stress	−0.25	0.01	−25.1	<.001	0.47	725	−0.26	0.01	−24.3	<.001	0.45	726
Sleep problems	−0.95	0.07	−14.2	<.001	0.24	725	−0.91	0.07	−12.9	<.001	0.21	726
Average pain intensity	−0.11	0.01	−10.6	<.001	0.13	723	−0.10	0.01	−9.9	<.001	0.12	724
Pain interference†	−0.49	0.03	−16.2	<.001	0.30	603	−0.48	0.03	−14.9	<.001	0.27	604
Depression	−0.33	0.02	−21.4	<.001	0.39	725	−0.34	0.02	−20.6	<.001	0.37	726
Future (8mo)												
SF-8 Mental	0.52	0.04	14.1	<.001	0.23	668	0.53	0.04	14.0	<.001	0.23	669
SF-8 Physical	0.57	0.03	17.2	<.001	0.36	668	0.57	0.04	16.3	<.001	0.33	669
Total fatigue	−1.18	0.05	−21.5	<.001	0.44	669	−1.21	0.06	−20.9	<.001	0.43	670
Perceived stress	−0.62	0.03	−20.1	<.001	0.38	670	−0.64	0.03	−19.9	<.001	0.38	671
Pain interference†	−0.40	0.03	−12.2	<.001	0.21	557	−0.39	0.04	−11.4	<.001	0.19	558
Depression	−0.32	0.02	−18.0	<.001	0.33	668	−0.32	0.02	−17.3	<.001	0.32	669

*All analyses adjusted for age and sex.
†Pain Interference Questionnaire completed only in those persons who indicated pain greater than zero.

Short Form

We constructed 6-item and 10-item short forms and scored participants using both forms. The difference between the 10- and 6-item forms was negligible, and the correlation between the scores was over .98. As a result, we decided to proceed with the 6-item short form. The 6 items were selected from the 17 items to achieve adequate measurement precision across the self-efficacy continuum. Items with higher discrimination parameters were given a preference while paying attention to the content of the items in order for the short form to provide good content coverage. The 6-item T score explains 95% of the variance in the 17-item T score, and the correlation between the short form (6 items) and the long form (17 items) was .97. The 6 items comprising the short form are specifically marked in table 2, and a formatted, ready-to-use downloadable PDF is available at http://uwcorr.washington.edu/publications.

DISCUSSION

The results suggested that the UW-SES functions similarly in both populations (MS/SCI) and both sexes in spite of the different distribution of men and women typical for the 2 disease conditions. The UW-SES score was correlated with the theoretically related constructs in ways previously reported. Specifically, as in our study, negative moderate correlations were reported between self-efficacy and depression (−.56, −.52, −.61) and anxiety (−.49, −.50, −.58) by Airlie,[14] Rigby,[5] and Middleton[15] and colleagues, respectively. This suggests that people with higher depressive symptoms and anxiety report lower self-efficacy. Middleton[15] reported a small positive correlation (r=.04) between the FIM motor score (n=34) and the score on the Moorong Self-Efficacy Scale. In our study, the FIM score was not available and the correlation between the physical subscale of SF-8 and the UW-SES score was a lot stronger (.57), suggesting that people with better physical function report higher self-efficacy. Our results differ from the results by Horn et al[16] who found that having a high-school education or greater was associated with statistically significantly higher self-efficacy. The correlation between educational level and self-efficacy score in our study was relatively small although in the same direction as previously reported, suggesting that people with higher educational levels report higher self-efficacy. Previous studies did not report on the relationship between social functioning and self-efficacy. In this study, we found a strong positive relationship between social functioning and self-efficacy, suggesting that people who are more confident that they can manage their symptoms do more and as a result are happier with their social roles and social function. Miller[53] reported a moderate positive correlation between satisfaction with life and self-efficacy (r=.51, *P*<.001), a finding that is similar to ours.

The newly developed UW-SES can be administered as a long form (17 items) or a short form (6 items). The item parameters are available and can be programmed into computerized adaptive test software to be administered dynamically. The instrument is free for all noncommercial purposes and can be found at http://uwcorr.washington.edu/publications. Administering the items dynamically does not require a validation because both computer adaptive testing and fixed number instruments (ie, short forms) use the same item parameters. However, studies to examine how many items would typically be administered by computer adaptive testing to achieve adequate precision and comparing the scores to both long and short forms would be useful. Custom-made short forms can be developed and scored using the published IRT parameters. Custom short forms may be useful for measuring people with specific levels of self-efficacy (for instance, to measure people with low self-efficacy with greater precision) or to achieve a certain level of precision (for instance, by developing a short form with 10 items). Because the instrument was developed with IRT as an item bank, new items can be added. The scale may be applicable to other chronic conditions, and we are working to validate the scale in other populations. Future studies could develop specific norms (for instance, for people with different types of MS or with different level of injury in SCI) to help with the score interpretation for people with different disease severity and course.

Study Limitations and Future Directions

The data used in the development of the scale were collected at different time-points of a longitudinal survey. Future studies involving different samples of people with MS and SCI will be needed to provide further evidence of the validity of the score. No test-retest reliability studies have been conducted in this study to examine the stability of the score, and no investigation of DIF by race could be conducted because of low numbers of individuals from different ethnic backgrounds.

The scale is less precise in measuring people with high self-efficacy, and so it may benefit from adding items that measure well at the high levels of self-efficacy. Investigators interested in differentiating among people with high self-efficacy may need to develop and test items that measure with greater precision at the high end of the continuum. Our primary research interest is in identifying people with low self-efficacy and developing interventions to increase low self-efficacy. Therefore, at the moment we have no plans to improve the scale's precision at high levels of self-efficacy. Future studies should examine the predictive validity of the self-efficacy scores over a longer period of time than 8 months used in the preliminary validation studies, evaluate sensitivity to change in a context of a treatment expected to change self-efficacy, and develop estimates for clinically minimally important differences. Because the scale was developed with IRT, new items could also be developed to better measure higher levels of self-efficacy. Finally, the strong relationship between social functioning and self-efficacy warrants further study.

CONCLUSIONS

Overall, the results of the analyses suggested excellent psychometric properties including high reliability, low floor and ceiling effects, well-functioning response categories, and strong evidence of the validity of the IRT-based score. The DIF analyses suggested that the items function the same way in both populations, and the scores are driven by the level of self-efficacy rather than by a type of condition, age, or sex. Formatted long and short forms with the scoring instructions can be downloaded from http://uwcorr.washington.edu/publications.htm.

References

1. Bandura A. Social foundations of thought and action: a social cognitive theory. Englewood Cliffs: Prentice-Hall; 1986.
2. Tedman S, Thornton E, Baker G. Development of a scale to measure core beliefs and perceived self efficacy in adults with epilepsy. Seizure 1995;4:221-31.
3. Barnwell AM, Kavanagh DJ. Prediction of psychological adjustment to multiple sclerosis. Soc Sci Med 1997;45:411-8.
4. Lorig K, Chastain RL, Ung E, Shoor S, Holman HR. Development and evaluation of a scale to measure perceived self-efficacy in people with arthritis. Arthritis Rheum 1989;32:37-44.
5. Rigby SA, Domenech C, Thornton EW, Tedman S, Young CA. Development and validation of a self-efficacy measure for people with multiple sclerosis: the Multiple Sclerosis Self-efficacy Scale. Mult Scler 2003;9:73-81.
6. Hampton NZ. Subjective well-being among people with spinal cord injuries: the role of self-efficacy, perceived social support, and perceived health. Rehabil Counsel Bull 2004;48:31-7.
7. Fortinsky RH, Kercher K, Burant C. Measurement and correlates of family caregiver self-efficacy for managing dementia. Aging Ment Health 2002;6:153-60.
8. Altmaier EM, Russel DW, Kao CF, Lehmann TR, Weinstein JN. Role of self-efficacy in rehabilitation outcome among chronic low back pain patients. J Counsel Psychol 1993;40:335-9.
9. Regenold M, Sherman MF, Fenzel M. Getting back to work: self-efficacy as a predictor of employment outcome. Psychiatr Rehabil J 1999;22:361-7.
10. Schwarzer R, Boehmer S, Luszczynska A, Mohamed NE, Knoll N. Dispositional self-efficacy as a personal resource factor in coping after surgery. Pers Individ Dif 2005;39:807-18.
11. Hampton NZ. Self-efficacy and quality of life in people with spinal cord injuries in China. Rehabil Couns Bull 2000;43:66-74.
12. Marks R, Allegrante JP, Lorig K. A review and synthesis of research evidence for self-efficacy-enhancing interventions for reducing chronic disability: implications for health education practice (part II). Health Promot Pract 2005;6:148-56.
13. Sherer MMJ, Mercandante B, Prentice-Dunn S, Jacobs B, Rogers R. The Self-Efficacy Scale: construction and validation. Psychol Rep 1982;51:663-71.
14. Airlie J, Baker GA, Smith SJ, Young CA. Measuring the impact of multiple sclerosis on psychosocial functioning: the development of a new self-efficacy scale. Clin Rehabil 2001;15:259-65.
15. Middleton JW, Tate RL, Geraghty TJ. Self-efficacy and spinal cord injury: psychometric properties of a new scale. Rehabil Psych 2003;48:281-8.
16. Horn W, Yoels W, Wallace D, Macrina D, Wrigley M. Determinants of self-efficacy among persons with spinal cord injuries. Disabil Rehabil 1998;20:138-41.
17. Shnek ZM, Foley FW, LaRocca NG, et al. Helplessness, self-efficacy, cognitive distortions, and depression in multiple sclerosis and spinal cord injury. Ann Behav Med 1997;19:287-94.
18. Jerusalem M, Schwarzer R. Self-efficacy as a resource factor in stress performance. In: Schwarzer R, editor. Self-efficacy: thought control of action. Washington (DC): Hemisphere; 1992. p 195-213.
19. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program for patients with chronic disease. Eff Clin Pract 2001;4:256-62.
20. Bamer AM, Johnson KL, Amtmann D, Kraft GH. Prevalence of sleep problems in individuals with multiple sclerosis. Mult Scler 2008;14:1127-30.
21. Bamer AM, Johnson KL, Amtmann DA, Kraft GH. Beyond fatigue: assessing variables associated with sleep problems and use of sleep medications in multiple sclerosis. Clin Epidemiol 2010;2:99-106.
22. Baylor C, Yorkston K, Bamer A, Britton D, Amtmann D. Variables associated with communicative participation in people with multiple sclerosis: a regression analysis. Am J Speech Lang Pathol 2011;19:143-53.
23. Johnson KL, Bamer AM, Yorkston KM, Amtmann D. Use of cognitive aids and other assistive technology by individuals with multiple sclerosis. Disabil Rehabil Assist Technol 2009;4:1-8.
24. Jensen MP, Hirsh AT, Molton IR, Bamer AM. Sleep problems in individuals with spinal cord injury: frequency and age effects. Rehabil Psychol 2009;54:323-31.
25. Ware JE, Kosinski M, Dewey JE, Gandek B. A manual for users of the SF-8 Health Survey. Lincoln: QualityMetric Incorporated; 2001.
26. Fisk JD, Ritvo PG, Ross L, Haase DA, Marrie TJ, Schlech WF. Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. Clin Infect Dis 1994;18:S79-83.
27. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24:385-96.
28. Cohen S, Williamson G. Perceived stress in a probability sample of the U.S. In: Soacapam S, Oskamp S, editors. The social psychology of health: Claremont Symposium on Applied Social Psychology. Newbury Park: Sage; 1988. p 31-67.
29. Spritzer KL, Hays RD. MOS Sleep Scale: a manual for use and scoring, Version 1.0. Los Angeles: RAND; 2003.

30. Becker J, Schwartz C, Saris-Baglama RN, Kosinski M, Bjorner JB. Using item response theory (IRT) for developing and evaluating the Pain Impact Questionnaire (PIQ-6). Pain Med 2007;8: S129-44.
31. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16:606-13.
32. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999;282:1737-44.
33. Patient-Reported Outcomes Measurement Information System. PROMIS Scoring Guide: Version 1.0 short forms, profile short forms, computerized adaptive testing. January 2011. Available at: http://www.assessmentcenter.net/documents/PROMIS%20Scoring%20Manual-%20CATs,%20Profiles,%20Short%20Forms.pdf. Accessed April 2012.
34. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992;50:133-49.
35. Bentler P. Comparative fit indices in structural models. Psychol Bull 1990;107:238-46.
36. Hu LT, Bentler P. Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6:1-55.
37. Hu LT, Bentler PM. Evaluating model fit. In: Hoyle RH, editor. Structural equation modeling: concepts, issues and applications. Thousand Oaks: Sage; 1995. p 76-9.
38. McDonald R. Test theory: a unified treatment. Mahway: Erlbaum; 1999.
39. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing structural equation models. Newbury Park: Sage; 1993. p 136-72.
40. Kline RB. Principles and practice of structural equation modeling. New York: The Guilford Press; 1998.
41. West SG, Finch JF, Curran PJ. SEM with nonnormal variables. Thousand Oaks: Sage; 1995.
42. Cook KF, Kallen MA, Amtmann D. Having a fit: impact of number of items and distribution of data on traditional criteria for assessing IRT's unidimensionality assumption. Qual Life Res 2009;18:447-60.
43. Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. Qual Life Res 2007;16:19-31.
44. McDonald RP. The dimensionality of test and items. Br J Math Stat Psychol 1981;34:100-17.
45. Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement 1969; No. 17:1-100.
46. Orlando M, Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. Appl Psychol Meas 2000;24:50-64.
47. Orlando M, Thissen D. Further investigation of the performance of S-X2: an item fit index for use with dichotomous item response theory models. Appl Psychol Meas 2003;27:289-98.
48. Scott NW, Fayers PM, Aaronson NK, et al. A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. J Clin Epidemiol 2009;62:288-95.
49. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: Erlbaum; 1988.
50. Zumbo BD. A handbook on the theory and methods of differential item functioning (DIF): logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense; 1999.
51. Crane PK, Belle GV, Larson EB. Test bias in a cognitive test: differential item functioning in the CASI. Stat Med 2004;23:241-56.
52. Crane PK, Gibbons LE, Ocepek-Welikson K, et al. A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. Qual Life Res 2007;16:69-84.
53. Miller SM. The measurement of self-efficacy in persons with spinal cord injury: psychometric validation of the Moorong Self-Efficacy Scale. Disabil Rehabil 2009;31:988-93.

Suppliers

a. Mplus v5.21, Muthén & Muthén, 3463 Stoner Ave, Los Angeles, CA 90066.
b. Multilog v7, Scientific Software International, Inc, P.O. Box 4728, Skokie, IL 60077.
c. IRTFIT: SAS macro, Quality Metric, Inc., 24 Albion Rd, Lincoln, RI 02865.
d. IRTSCORE, L.L. Thurstone Psychometric Laboratory, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599.
e. Logistic Regression Differential Item Functioning (LORDIF): The Comprehensive R Archive Network: http://cran.r-project.org/web/packages/lordif/index.html.