American Psychological Association

Rehabilitation Psychology

ISSN: 0090-5550

2020, Vol. 65, No. 2, 145–157
http://dx.doi.org/10.1037/rep0000312

Development of a Resilience Item Bank and Short Forms

Dagmar Amtmann, Alyssa M. Bamer, Kevin N. Alschuler, Fraser D. Bocell, Dawn M. Ehde, Mark P. Jensen, Kurt Johnson, Mara B. Nery-Hurwit, Rana Salem, Arielle Silverman, and Amanda E. Smith
University of Washington

Alexandra L. Terrill
University of Utah

Ivan Molton
University of Washington

Purpose: The purpose of this study was to develop a publicly available, psychometrically sound item bank and short forms for measuring resilience in any population, but especially resilience in individuals with chronic medical conditions or long-term disability. ***Research Methods:*** A panel of 9 experts including disability researchers, clinical psychologists, and health outcomes researchers developed a definition of resilience that guided item development. The rigorous methodology used focus groups, cognitive interviews, and modern psychometric theory quantitative methods, including item response theory (IRT). Items were administered to a sample of people with chronic medical conditions commonly associated with disability (N = 1,457) and to a general population sample (N = 300) representative of the Unites States general population with respect to age, gender, race, and ethnicity. ***Results:*** The final item bank includes 28 items calibrated to IRT with the scores on a T-metric. A mean of 50 represents the mean resilience in the general population sample. Four and eight item short forms are available, and their scores are highly correlated with the item bank score ($r \geq$ .94). Reliability is excellent across most of the resilience continuum. Initial analyses provide strong support for validity of the score. ***Conclusions:*** The findings support reliability and validity of the University of Washington Resilience Scale (UWRS) for assessing resilience in any population, including individuals with chronic health conditions or disabilities. It can be administered using computerized adaptive testing or by short forms.

Impact and Implications

This study describes the development and psychometric properties of the first resilience measure using stakeholder engagement to define the construct as well as modern psychometric methods, including item response theory (IRT). The UWRS has significant potential to advance both practice and research through its range of uses. It offers flexibility to use computerized adaptive testing or short forms, reducing respondent burden. It is suitable for assessing resilience in the general population as well as people with chronic health conditions or disability. The UWRS is also freely available, which facilitates its use in both research and clinical practice. Use of this psychometrically sound, person-centered, brief, flexible, and free resilience instrument will facilitate observational and intervention research, as well as research comparing resilience across populations and studies.

Keywords: resilience, item response theory, measurement, scale development

Editor's Note. Stephen Wegener served as handling editor.—DE

This article was published Online First February 10, 2020.

Dagmar Amtmann, Alyssa M. Bamer, Kevin N. Alschuler, Fraser D. Bocell, Dawn M. Ehde, Mark P. Jensen, Kurt Johnson, Mara B. Nery-Hurwit, Rana Salem, Arielle Silverman, and Amanda E. Smith, Department of Rehabilitation Medicine, University of Washington; Alexandra L. Terrill, Department of Occupational and Recreational Therapies, University of Utah; Ivan Molton, Department of Rehabilitation Medicine, University of Washington.

Amanda E. Smith is now at Institute for Health Metrics and Evaluation, University of Washington.

Authors would like to acknowledge conceptual guidance provided by Ken Pakenham in the process of development of the definition of resilience. We certify that we have complied with the APA ethical principles regarding research with human participants and/or care and use of animals in the conduct of the research presented in this article.

The contents of this article were developed under grants from the Administration for Community Living, National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR; formerly NIDRR), Grants H133B080024 and 90RT5023-01-00. However, those contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government. In addition, NIDILRR had no involvement in the data collection, analysis, interpretation, manuscript preparation, or decision to submit the article for publication. Preliminary results of this study were presented at the International Society for Quality of Life Research annual meetings in 2016 (Copenhagen, Denmark) and 2017 (Philadelphia, Pennsylvania).

Correspondence concerning this article should be addressed to Dagmar Amtmann, PhD, Department of Rehabilitation Medicine, University of Washington, Box 354237, 4907 25th Avenue North East, Seattle, WA 98105. E-mail: dagmara@u.washington.edu

Introduction

Although there is no universally agreed upon definition of resilience, (Johnston et al., 2015; Stewart & Yuen, 2011) it is broadly defined as a dynamic process by which individuals demonstrate positive psychosocial adaptation in the context of adverse events and/or ongoing stressors or adversity (Black & Dorstyn, 2015; Earvolino-Ramirez, 2007; Southwick, Bonanno, Masten, Panter-Brick, & Yehuda, 2014; Terrill, MacKenzie, Einerson, & Reblin, 2019). The American Psychological Association (APA) defines resilience as "the process of adapting well in the face of adversity, trauma, tragedy, threats or even significant sources of stress" (APA, 2014). Resilience has been studied in the context of a diverse array of adversities, including bereavement, trauma, childhood adversity, natural disasters, and life-threatening or chronic medical conditions. Individuals characterized as resilient in a particular context typically maintain and/or recover their psychological, physiological, and social equilibrium after experiencing a stressor or adverse event. Importantly, resilience also includes the process of sustaining well-being and progressing toward life goals in the context of chronic stressors, including the challenges associated with chronic health conditions (Zautra, 2009). Although certain individual traits, genetics, and early life experiences may predispose some individuals toward demonstrating resilience (Bonanno, 2004; Steptoe, Dockray, & Wardle, 2009), resilience is not viewed as a fixed personality trait (Johnston et al., 2015). Rather, contemporary conceptualizations emphasize the dynamic nature of resilience in which risk and protective factors are thought to play important, intersecting roles (Black & Dorstyn, 2015). As such, resilience may be viewed as a constant exchange between a person's internal capacities and external resources within their broader interpersonal or social environments in response to a stressor or adversity (Terrill et al., 2019) and varies with context and across the life span (Southwick et al., 2014).

Resilience has been associated with better quality of life for those with disabilities (Manning, Carr, & Kail, 2016; Moore et al., 2015; Silverman, Molton, Alschuler, Ehde, & Jensen, 2015). Moreover, recent studies that have examined resilience in a large sample of adults aging with long-standing physical disability found that those who are more resilient have significantly fewer depressive symptoms, higher quality of life—in spite of relatively high levels of pain and/or fatigue—and better social participation and satisfaction with social roles (Battalio et al., 2017; Silverman et al., 2015; Terrill et al., 2016). Longitudinal research in a sample of adults with multiple sclerosis (MS) also found that increases in resilience were associated with improved mood, improved sleep quality, and increased physical activity; conversely, decreases in resilience over time were associated with increased fatigue severity and depressive symptoms (Koelmel, Hughes, Alschuler, & Ehde, 2017). Because higher resilience is associated with better health, social functioning, and quality of life in rehabilitation populations, it has been identified as a therapeutic target (Alschuler, Arewasikporn, Nelson, Molton, & Ehde, 2018; Black & Dorstyn, 2015; Leppin et al., 2014; Loprinzi, Prasad, Schroeder, & Sood, 2011; Petriwskyj, Parker, O'Dwyer, Moyle, & Nucifora, 2016; Reivich, Seligman, & McBride, 2011).

While much has been learned about resilience, because of the heterogeneity in the operationalization of resilience, research on the construct continues to be limited by the lack of psychometrically robust resilience measures, resulting in difficulties comparing results across populations and studies. In a 2011 review of resilience measures, Windle, Bennett, & Noyes (2011) found that there is no one "gold standard" measure of resilience, suggested the evidence to support any one measure was moderate at best, and indicated that all of the 15 measures they identified required additional evaluation of validity. The complexity of the dynamic resilience process is also not adequately captured by current measures (Johnston et al., 2015). Further, none of the currently available measures provide a "normal" or "beneficial" level of resilience, making it difficult to establish an outcome for targeted intervention. To evaluate interventions that promote resilience, universally applicable (rather than disease specific), reliable, and valid measures of resilience are also needed.

It is notable that none of the previously reviewed universally applicable resilience measures were developed using modern psychometric methods, that is, item response theory (IRT). In fact, we were only able to identify one recently developed disease specific measure, Spinal Cord Injury—Quality of Life (SCI-QOL) Resilience that used IRT (Victorson et al., 2015). Compared with classical test theory, IRT methodology can significantly improve psychometric properties of multiitem scales. IRT-based instruments provide more flexibility and potentially significant reductions in assessment time, especially when computer adaptive testing (CAT) is used for administration. CAT uses a mathematical algorithm to tailor items to the respondent based on responses to previously administered items and results in a reliable score while administering few items (De Ayala, 2009). IRT based instruments provide scores on the same metric regardless of the mode of administration or what items were administered. This flexibility facilitates the use of outcome measures in research and clinical practice. The utility of IRT-based instruments can further be enhanced by centering the metric on the mean that represents the U.S. general population. This offers an intuitive interpretation of the scores and allows for direct comparisons across populations and studies. As a result, IRT-based instruments have become increasingly popular in health assessment, quality-of-life research, and clinical research (Cella et al., 2012; Gershon et al., 2012; Tulsky et al., 2015, 2016).

Given the lack of suitable available measures, we sought to develop a universally applicable, reliable, and valid measure of resilience called the University of Washington Resilience Scale (UWRS). The aim of this study was to develop a publicly available, psychometrically sound IRT-based item bank and short forms suitable for measuring resilience in any population, but specifically relevant to resilience in the context of chronic medical conditions or disability.

Method

Development of the resilience item bank followed the methodology used by the Patient Reported Outcomes Measurement Information System (PROMIS) network (Reeve et al., 2007) and by other national measurement initiatives (Gershon et al., 2012, 2013). In addition to using modern measurement theory, the standard procedures involve extensive feedback from the target audiences (e.g., people with health conditions or disabilities, and their health care providers) in the development of the

construct definition and to make sure items are meaningful and relevant to the target audiences. The following subsections report the specific steps used to create a resilience item bank including development of a domain definition, expert panel review, testing of items in cognitive interviews, administration of the items to a sample of individuals with disabilities as well as a general population sample, and results of the IRT analyses. All relevant research procedures were reviewed and approved by the University of Washington Human Subjects Institutional Review Board.

Participants

Focus group participants. Four focus groups were conducted, two with middle-aged (36–62 years) individuals with MS (one with men, $n = 6$, and one with women, $n = 6$), one for partners of individuals with MS ($n = 11$) and one with community stakeholders serving people with MS ($n = 9$). Participants were at least 18 years of age and were recruited from a variety of sources including clinician referral and local MS support groups, as described in detail in Silverman, Verrall, Alschuler, Smith, and Ehde (2017). The average age of the participants with MS was 52 years, and they had been living with MS for an average of 14 years.

Cognitive interview participants. The items developed by the research team based on focus group results and literature review were tested in cognitive interviews with people at least 18 years of age, able to read and speak English, and who have one of four conditions commonly associated with disability (MS, muscular dystrophy [MD], spinal cord injury [SCI], or postpolio syndrome [PPS]). Participants from previous research studies (for examples see (Amtmann, Cook, Johnson, & Cella, 2011; Ehde et al., 2015; Molton et al., 2014)) were invited to participate, and all 11 participants invited completed the cognitive interview. The average age of the cognitive interview participants was 61 years, four (36%) were men, six had MS, two had PPS and SCI, and one had MD.

Large sample of people with chronic medical conditions commonly associated with disability. The candidate items were administered to all individuals who participated in an ongoing national longitudinal survey study of adults aging with a physical disability at the University of Washington (Molton et al., 2014; Smith, Molton, & Jensen, 2016; hereafter called the "clinical sample" because it includes people with specific diagnoses). Participants were required to self-report one of four long-term conditions associated with disability, including MS, SCI, PPS, MD, be 18 years of age or older, and read, write, and understand English. A total of 1,573 individuals from one time point of the longitudinal survey were invited to participate in the calibration survey study, of which 1,457 (93% response rate) returned the survey, and 1,441 provided responses to resilience items. The clinical sample had a mean age of 63 years, 65% were women, and 90% were non-Hispanic White (see Table 1). The responses from the 1,441 individuals were used to calibrate the final items to IRT.

U.S. general population sample. A U.S. general population community-based sample of adults living in the U.S. was recruited using Toluna (http://www.toluna-group.com/), an online panel and survey-technology provider. Toluna maintains a large database of people who agreed to participate in online surveys (Toluna, 2017). Participants were matched to the 2010 U.S. Census on gender (49% male, 51% female), age (49% 18–44, 34% 45–64, 17% 65+ years), race (75% White, 12% Black, 4% Asian, 9% other or more than one race), and ethnicity (16% Hispanic; Howden & Meyer, 2011; Humes, Jones, & Ramirez, 2011). Outside of these demographic characteristics, the participants were not selectively recruited to be healthy, and as a result, respondents may have had different health conditions as they occur in the community. A total of 538 individuals were invited to participate in the U.S. general population sample and 378 completed the survey (70% response rate). Of these, 78 were dropped because of invalid responses (e.g., inconsistent responses, long strings of the same responses, time too short to read questions), resulting in a final U.S. general population sample of 300 individuals (56% response rate) matched to the U.S. census on gender, age, race, and ethnicity. The mean age of the U.S. general population sample was 49 years, 51% were women, and 63% were white and non-Hispanic (see Table 1). The mean resilience of the U.S. general population sample was set as the mean of the resilience metric (i.e., $M = 50$ represents the mean resilience of the U.S. general population).

Procedure

Focus groups. A total of four focus groups were conducted with the purpose of exploring what resilience means, how to define it, and to identify the most important facets of resilience. The semistructured focus groups were facilitated by two experienced moderators. The focus group discussions were transcribed and analyzed by two researchers who identified themes and subthemes. A detailed discussion of the methods and process used is available in Silverman et al. (2017).

Expert panel. A panel of nine experts including clinical psychologists, health outcomes researchers with training in measurement, and disability researchers met on multiple occasions over several months to discuss the definition and important facets of resilience. Several expert panel members also facilitated or participated in the focus groups. Results of the focus groups informed the definition of resilience, including important subdomains. Items were generated by expert panel members individually using the domain definition and the time frame, type of item stems (questions vs. statements), and response options agreed on. The draft items proposed by the individual expert panel members were refined or deleted by consensus, taking into account content coverage, consistency of wording, and clarity.

Cognitive interviews. Preliminary items were tested in structured cognitive interviews (Willis, 2005) to ensure that items were understandable, relevant and meaningful to target audiences. Each item was reviewed by at least five individuals. The cognitive interviewing process involved asking respondents to respond to all resilience items, and immediately after responding, asked how respondents arrived at their answer as well as additional probing questions about clarity, appropriateness, and meaningfulness of the question. Participants were also asked to provide feedback on the response options and instructions. Items were modified and/or new items generated based on interview feedback, with subsequently generated items undergoing the same cognitive interview process until all items were functioning well and covered all areas participants considered important. Cognitive interviewees were paid $25 for their time.

Table 1
University of Washington Resilience Scale (UWRS) Participant Characteristics for the Calibration Sample of People With Chronic Medical Conditions (i.e. Clinical Sample) and the U.S. General Population Sample

Variable	Clinical sample (N = 1,457) $M \pm SD$ n (%)	U.S. general population sample (N = 300) $M \pm SD$ n (%)
Age (years)	62.5 ± 12.1	47.5 ± 16.6
Disability group		
Spinal cord injury	349 (24.0)	N/A
Postpolio syndrome	383 (26.3)	
Muscular dystrophy	274 (18.8)	
Multiple sclerosis	451 (31.0)	
Sex		
Women	946 (64.9)	154 (51.3)
Men	511 (35.1)	146 (48.7)
Marital status		
Married/living with partner	895 (61.5)	208 (69.3)
Separated/divorced	281 (19.3)	30 (10.0)
Never married	154 (10.6)	41 (13.7)
Widowed	126 (8.7)	21 (7.0)
Education		
<High school	25 (1.7)	4 (1.3)
High school graduate/GED	170 (11.7)	39 (13.0)
Some college/vocational degree	429 (29.5)	70 (23.3)
Bachelor degree	433 (29.8)	115 (38.3)
Professional/graduate degree	398 (27.4)	72 (24.0)
Household income		
<$20,000	179 (14.7)	46 (15.3)
$20,000–$51,999	430 (35.3)	84 (28.0)
$52,000–$99,999	331 (27.1)	99 (33.0)
>$100,000	279 (22.9)	71 (23.7)
Race/ethnicity		
Non-Hispanic White	1,302 (90.0)	189 (63.0)
Non-Hispanic Black	58 (4.0)	34 (11.3)
Non-Hispanic other race	56 (3.9)	27 (9.0)
Hispanic White or other race	30 (2.1)	50 (16.7)
Employment		
Full-time employment or student	182 (12.5)	134 (44.7)
Part-time employment	141 (9.7)	38 (12.7)
Unemployed—retired, disability, or homemaker	1,134 (77.8)	128 (42.7)

Note. GED = passed General Educational Development test; *SD* = standard deviation.

Large scale administration to clinical sample. Items modified to reflect the feedback from cognitive interviews were administered to a clinical sample in a paper survey which was mailed between October, 2015 and March, 2016. Surveys were scanned using the Cardiff TeleForm Optical Mark Recognition software program (Cardiff, 2015) and participants were paid $25 for their participation.

Measures. The survey included demographic (e.g., race, ethnicity, income, education) and medical condition related questions (e.g., disease duration, physical function, provider access) as well as multiple health and quality of life outcome measures. Measures most relevant to this study included the Connor-Davidson Resilience Scale 10-item short form (Campbell-Sills & Stein, 2007), the Flourishing Scale (Diener et al., 2010), the Positive and Negative Affect Schedule (PANAS) short forms (Mackinnon et al., 1999), the PROMIS Depression short form 6a (Pilkonis et al., 2011), and PROMIS Physical Function with Mobility Aid short form (Rose et al., 2014) scale. The Connor-Davidson Resilience scale is commonly used to measure resilience, and was developed for clinical practice as a measure of stress coping ability and has been reported to have adequate construct validity and internal consistency (G. Windle, 2010). Items are summed and higher scores indicate more resilience. The Flourishing Scale is a measure of psychological well-being based on humanistic and eudaimonic well-being theories (Diener et al., 2010), and has been found to have acceptable psychometric properties in a variety of study populations, including those with low mental health (Schotanus-Dijkstra et al., 2016). The Flourishing Scale includes eight items which are summed for a total score, and higher scores indicate greater well-being. The PANAS scale is one of the most widely used scales for evaluating subjective well-being, and includes two subscales to measure positive and negative affect. For both scales items are summed and higher scores indicate either more positive affect or more negative affect. The PANAS short form has been found to have good factorial validity and invariance across demographic variables in a community sample (Mackinnon et al., 1999). The PROMIS Depression and Physical Function scales were developed using modern psychometric methods and have been shown to have strong

validity and reliability in diverse clinical samples (Schalet, Hays, et al., 2016; Schalet, Pilkonis, et al., 2016). Scores on both measures are centered on the U.S. general population and higher scores indicate more depression or better physical function, respectively (Pilkonis et al., 2011; Rose et al., 2014).

Centering the score on the U.S. general population mean. Items were administered to a U.S. general population community sample using Toluna (http://www.toluna-group.com/). Toluna delivered study invitations (n = 538) via e-mail, mobile text, or through an application, with a direct link to the survey provided by the authors. The online survey was administered through REDCap (Research Electronic Data Capture; Harris et al., 2009) hosted at the University of Washington, and the survey included resilience items and demographics including sex, age, race/ethnicity, income, marital status, education, and employment status. Participant responses (n = 378) were carefully reviewed and records were dropped (n = 78) if responses were inconsistent (e.g., same responses to negatively and positively worded items or age at screening different from age in survey response) or invalid (e.g., average response time for all items was ≤2 min). Participants received a small incentive from Toluna upon completing the survey (Toluna awards survey participants points which may be redeemed for products when a sufficient number of points are accrued).

Data Analyses

Sample means and standard deviations were calculated for all measures and grouped by gender, medical diagnosis groups, and age groups in both the clinical and U.S. general population samples. General descriptive statistics for both samples were also calculated for demographics (e.g., age, race, ethnicity, education).

IRT analyses. Using data from the clinical sample, analyses were completed to examine IRT assumptions (unidimensionality and local independence), to calibrate items to a graded response IRT-model (Samejima, 1996), to examine differential item functioning (DIF), and to evaluate the reliability of the scores based on full bank and short forms as described below. One factor confirmatory factor analysis (CFA) was used to examine whether the responses to the items are sufficiently unidimensional using the WLSMV estimator in Mplus 7.2 (Muthén & Muthén, 1998–2012). A comparative fit index (CFI) of 0.90 or higher (Hu & Bentler, 1999; Reise, Widaman, & Pugh, 1993) was interpreted as evidence of sufficient unidimensionality. Local independence violations were evaluated using residual correlations from the CFA. Residual correlations greater than 0.20 were interpreted as evidence of local dependence (LD; Kim, De Ayala, Ferdous, & Nering, 2011). The CFA is used to check that the unidimensionality assumptions of the IRT are met, before proceeding with fitting the IRT model.

Items were calibrated to a graded response IRT model using the *mirt* package in R (Chalmers, 2012). Item fit was evaluated using Orlando and Thissen's (2003) S-χ^2 with the IRTFIT macro (Bjorner, Smith, Stone, & Sun, 2007) in SAS Version 9.3 (SAS Institute, Inc., 2012–2013). A S-χ^2 p value of >.01 was considered adequate to retain the item in the bank. Both *mirt* and IRTFIT utilize list wise deletion of missing data resulting in loss of data from participants with incomplete responses. Items displaying LD or poor fit were removed and the models rerun until all items met the criteria for local independence and had acceptable fit statistics.

Individual item response curves generated after fitting the graded response model were inspected to evaluate the appropriateness and functioning of the response options. In addition, the test information function graphs generated by the IRT analyses were used to examine the reliability of the scores across the levels of resilience. Information functions were converted to reliability and plotted together with a histogram of participant responses and plotted along the T score continuum. Scale information of 5 in the IRT framework corresponds to reliability of 0.8 in the classical test theory framework and information of 10 corresponds to reliability of 0.9 (Thissen, 2000). In general, reliability of 0.8 is considered sufficient for group comparisons while 0.9 is required for individual comparisons (Hahn et al., 2007; Nunnally & Bernstein, 1994). The score ranges with reliability greater than 0.8 or 0.9 (depending on the desired reliability) were calculated for the full bank and short forms.

DIF analyses were completed using *lordif* (Choi, Gibbons, & Crane, 2011) in R (R Core Team, 2016) after fitting items to the IRT model. DIF by sex (male vs. female), age (<60 vs. ≥60 years), diagnosis (PPS, MS, MD, SCI), and education (less than college degree vs. college or professional degree) was evaluated in the clinical sample. Presence of DIF was also evaluated by comparing the responses of the clinical sample and U.S. general population sample. Items were considered to have statistically significant DIF when a change in pseudo R^2 statistic of 0.13 (Zumbo, 1999) or a 5% change in β coefficients (Crane et al., 2007) was observed. Items with significant DIF were dropped from the bank. After dropping items with LD, misfit to the IRT model, or DIF, final item parameter estimates were generated using data from the clinical sample and centered on the mean of the U.S. general population sample (i.e., a score of 50 is the mean of the general population sample).

Short forms. Multiple considerations were taken into account when selecting items for inclusion in the short forms. Content of the items, as well as item parameters (i.e., difficulty and discrimination), functioning of response categories, and readability of individual items were considered. Readability was evaluated by calculating the Flesch-Kincaid Grade Level score (Kincaid, Fishburne, Rogers, & Chissom, 1975) using Microsoft Word, 2016 (Microsoft Corporation, 2016).

A four-item and eight-item short form were constructed to accommodate different length of administration needs and all four items on the four-item form are included on the eight-item version. Correlations between short form scores and scores based on all the items in the item bank were examined to evaluate the correspondence. Test information functions for the short forms were plotted to examine reliability and effective range of measurement of the short forms. Conversion tables for converting the summary score to IRT based T score for both short forms were generated using the program IRTSCORE (Flora & Thissen, 2002).

Validity. The construct validity of the UWRS was examined by calculating Pearson's correlation with existing instruments measuring the same or similar constructs. We hypothesized a strong correlation ($\sim r > .7$) with the Connor-Davidson Resilience scale, moderate to strong positive correlations ($\sim r > .5$) with the Flourishing Scale and PANAS positive affect; moderate to strong negative correlations ($\sim r > -0.5$) with PANAS negative affect

and PROMIS Depression scales; and a weak correlation ($\sim r < .3$) with PROMIS physical function.

Results

Focus Groups

Results of the focus groups suggested the following facilitators of resilience: psychological coping (e.g., flexibility, humor), social connections, life meaning, planning skills, and engagement in physical wellness activities (e.g., exercise, energy management). Barriers to resilience included resilience depletion (feeling "burned out"), negative thoughts and feelings, social barriers, social stigma, and physical fatigue. These themes helped guide the development of the domain definition and informed many of the candidate items. The results of the focus groups are described in detail in Silverman et al. (2017).

Definition of Resilience

Based on the literature and the results of the focus groups, the panel initially conceptualized resilience varying along three dimensions. These were: (a) *restoration versus maintenance* (whether a person returns to baseline, or maintained quality of life, engagement in activities, and well-being in the face of adversity; see Zautra, 2009); (b) *internal versus external* (reflecting resilience in terms of internal states, like positive affect and positive outlook, as opposed to behavioral resilience, i.e., efforts to continue toward valued goals in the face of adversity); and (c) *acute versus chronic* (to reflect the nature of the stressor). Although the nature of the stressor is not considered in most published definitions of resilience, the panel thought that chronicity was important to consider, given that many people (including individuals with disability) may face a large number of chronic, low-level hassles associated with health self-management, stigma, and restrictions in daily activities.

This initial conceptualization generated a $2 \times 2 \times 2$ grid. However, after further consideration, the panel elected to collapse *restoration versus maintenance* and *acute versus chronic* dimensions because the two domains overlap to a high degree, in that an individual is much more likely to manifest resilience as maintenance of affect/outlook or behavior in the face of a chronic stressor (as opposed to an acute one). Conversely, the concept of "maintenance" of affect/outlook or behavior is unlikely in the face of a high intensity, acute stressor. The panel therefore agreed that while a distinction between restoration/maintenance and chronicity may exist conceptually, items generated for these domains would not be different enough from one another to justify the additional category.

We therefore proceeded with a 2×2 model of resilience, reflecting resilience responses that could reflect restoration versus maintenance and internal versus external. In this approach, a person could demonstrate "resilience" if, in the presence of a new stressor or setback, their emotional well-being dipped and then returned to baseline, or if they reengaged to work toward a valued goal. They would also be considered resilient if in the face of a chronic stressor, they were able to maintain their engagement in valued activities, without a change from baseline, or maintain their affect/outlook. This multifactorial conceptualization was designed to include the range of ways in which people may demonstrate resilience to stressors in specific contexts, without overemphasizing either positive outlook or task persistence alone. The panel noted that most previous definitions have focused on essentially one square in the grid (i.e., restoration of positive affect/outlook after an acute stressor), and that this approach did not capture the reality of resilience in the face of chronic health-related problems, stigma, and other stressors for people facing an ongoing stressor. Importantly, this definition of resilience does not imply the absence of negative emotions, which may occur simultaneously with resilience (Zautra, 2009).

In summary, the panel defined resilience as *the capacity to bounce back from and/or maintain function in the face of adversity*. The panel felt the following points were also important regarding the definition of resilience: (a) resilience is generally stable across situations, but demonstrates context dependent variability and malleability; (b) resilience is activated by the presence of adversity; (c) resilient responses can include both internal states such as one's outlook as well as observable behaviors; (d) resilience can exist in the presence of a range of severity of stressors, including acute/major (e.g., trauma), acute/minor (e.g., goal setback), chronic/minor (e.g., daily hassles) and chronic/major (e.g., stigma, disability); (e) resilience may involve a return to prestressor baseline of function (i.e., "bouncing back") or maintenance of baseline function despite the presence of a stressor; (f) resilience may manifest differently based on the type and severity of stressor; and (g) resilience is related to, but distinct from related constructs including optimism, self-efficacy, hardiness, and posttraumatic growth. Forty-seven items were generated by the expert panel based on this definition.

Cognitive Interviews

The cognitive interview participants were then asked to rate how much each of the 47 resilience statements describes them. Five response options ranging from *not at all* to *very much* (DeWalt, Rothrock, Yount, & Stone, 2007) were provided. Due to the relatively stable nature of resilience, the expert panel decided the measure should ask about "right now" rather than use a longer time frame, such as in the past 7 days. Eleven items were identified as problematic during the cognitive interview process and dropped from the bank, resulting in 36 items that were administered to the large-scale calibration clinical sample. Examples of reasons items were dropped included phrases that were not commonplace or easily translated ("maintain an even keel" and "take things in stride"), double negatives that were confusing to answer ("When stressful events happen to me, I am not usually bothered by them"), or content taken literally rather than metaphorically ("When I get knocked down, I get up again").

Clinical and U.S. General Population Samples

IRT analyses. Two items of 36 were eliminated due to LD and six demonstrated significant misfit to the graded response model. Items in the final item bank ($n = 28$) were unidimensional (CFI = 0.95), did not show statistically meaningful levels of LD, and fit well the IRT model. Review of the item response curves generated by the IRT analyses indicated that the five response options functioned adequately for all items.

The 28 calibrated items were evaluated for DIF for the following categories: sex (male, $n = 509$ vs. female, $n = 933$), age (<60, $n = 545$ vs. ≥60 years, $n = 897$), education (<college degree, $n = 621$ vs. college or professional degree, $n = 821$), diagnosis group (SCI, $n = 347$; PPS, $n = 379$; MD, $n = 271$; MS, $n = 445$), and clinical sample versus U.S. general population sample (clinical sample, $n = 1{,}457$; U.S. general population, $n = 300$). No items were flagged for DIF using either the McFadden's pseudo R^2 criteria or the 5% change in β coefficients criteria. Therefore, no items were eliminated from the bank due to DIF.

Item slopes for the final 28 item set ranged from 1.15–4.02 and thresholds ranged from −3.53 to 1.95. Item text for all items are shown in Table 2 and item parameters are available upon request (e-mail uwcorr@uw.edu or corresponding author). The reliability of the full bank was high for a large majority of participants (see Figure 1), and was above 0.9 between a T score of 20 and 71, and above 0.8 between 20 and 75 (inclusively). The UWRS scores are centered on the U.S. general population sample such that a mean of 50 represents the mean resilience in the U.S. general population. The clinical sample reported slightly higher resilience (mean: 50.7) than the U.S. general population sample

Table 2
University of Washington Resilience Scale (UWRS) Items, Short Forms, and Item Fit

Item ID	Item stem	Short form inclusion	Bin	Item fit	
				χ^2	p-value
		4 = four-item 8 = eight-item	1 = recover outlook 2 = maintain outlook 3 = recover persistence 4 = maintain adaptive response		
resil4	I maintain a positive outlook even in bad circumstances.	4, 8	2	112.4	0.72
resil9	I can stay focused on my goals even when bad things happen in my life.		2 or 4	95.1	0.89
resil10	When something happens that makes me feel stressed, I usually calm down quickly.	4, 8	1	119.0	0.41
resil12	When bad things happen to me, I bounce back quickly.		1 or 3	74.5	0.99
resil15	When things get stressful, I keep focused on my values and goals.		2 or 4	107.2	0.58
resil16	When I get stressed out, I'm usually able to get back to my old self in no time.		1	142.0	0.13
resil22	When faced with challenges, I keep working toward a solution.		4	131.1	0.03
resil32	I maintain a clear head when something stressful happens.		2	103.8	0.80
resil35	I keep things in perspective when something stressful happens.		2	86.6	0.94
resil36	When I experience a set-back, I keep moving forward.	8	3	84.1	0.93
resil37	Even when I experience a set-back, I keep moving toward my goals.		3	119.3	0.16
resil38	When I experience a set-back, I know I am able to get back to my usual self.		1	90.7	0.89
resil42	Although I feel bad sometimes, I usually bounce right back.	8	1	92.9	0.86
resil43	When something really bad happens to me, I usually feel bad for a long time. (R)		1	157.2	0.09
resil45	It does not take me long to feel fine again after feeling stressed.		1	129.5	0.18
resil49	During stressful times, I am usually calm and relaxed.	8	2	128.8	0.59
resil55	I stay calm no matter what is thrown at me.		2	121.7	0.73
resil61	I bounce back when difficulties come my way.		1 or 3	76.5	0.95
resil68	I do what is important to me, even when stressed.	8	4	110.9	0.73
resil69	I take care of what is important to me when facing stress or adversity.		4	97.8	0.75
resil71	When something stressful happens, I keep going.	4, 8	4	118.0	0.18
resil72	I get back to my priorities after stressful events.		3	79.1	0.98
resil74	When things go wrong in my life, I can pick myself up and start again.	4, 8	3	93.8	0.71
resil76	People who know me say I can maintain a level head when things get stressful.		2	113.6	0.81
resil77	I am easily thrown off by everyday hassles. (R)		2	125.7	0.39
resil79	I tend to give up easily when I face set-backs. (R)		3 or 4	126.7	0.25
resil81	I stay focused on things that are important to me while under pressure.		2	121.3	0.28
resil82	I move forward despite adversity.		4	101.9	0.79

Note. (R) indicates item to be reverse coded when administered. Except for those marked as reverse coded, response options for all items are 1 = *not at all*, 2 = *a little bit*, 3 = *somewhat*, 4 = *quite a bit*, 5 = *very much*.

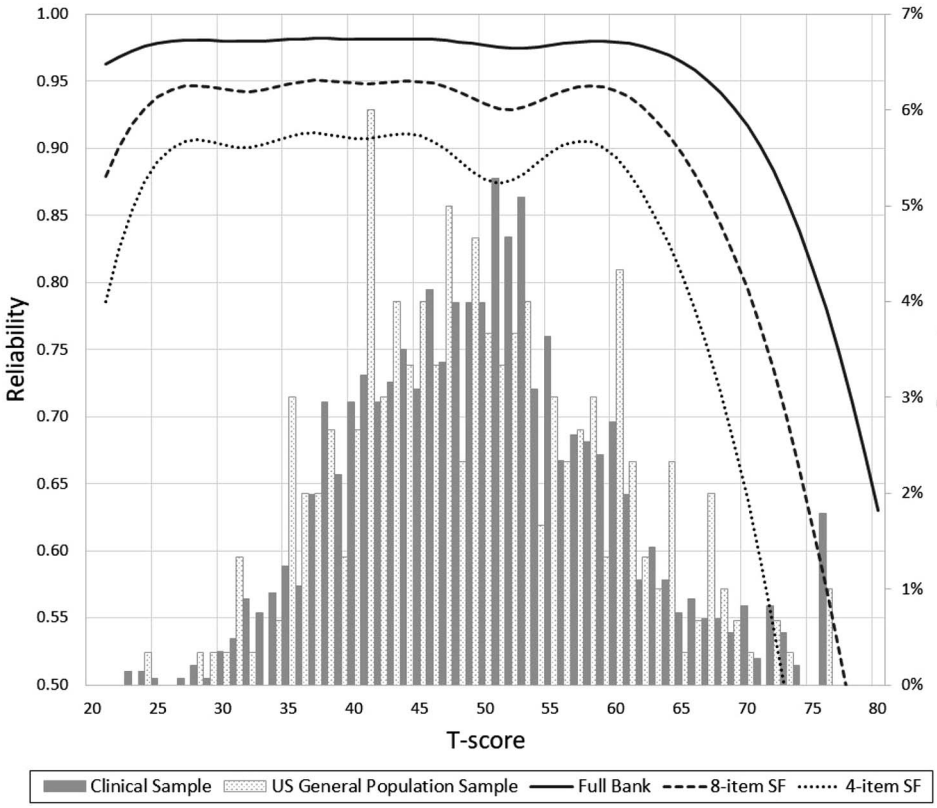

Figure 1. Information and reliability of the resilience item bank and eight- and four-item short forms (SFs) compared with the distribution of resilience *T* scores in the clinical sample and U.S. general population sample. Information in IRT corresponds to reliability in classical test theory. Information of 10 corresponds to reliability of 0.9. The information functions plotted on the graph show the reliability of the full item bank or short forms. Reliability varies at different levels of resilience, and is highest for the full bank between 20 and 71, 22 and 64 for the eight-item short form, and 26–46 and 55–59 for the four-item short form. The bars represent the frequency distribution of resilience scores in the two different study samples across the continuum of resilience.

(mean: 50.0). Resilience scores by gender, medical diagnosis groups, and age groups are presented in Table 3.

Short forms. An eight- and a four-item fixed length short form were developed for convenience when CAT is not feasible or desirable. Items for the short forms were selected by the authors based on their content, item parameters (i.e., their reliability at different resilience levels and ability to discriminate between people with similar resilience levels), functioning of response categories, and readability of individual items were considered. The four-item short form consists of a subset of items from the eight-item short form (short form items labeled in Table 2). Correlations between the short forms and the full bank were 0.97 for the eight-item short form, 0.94 for the four-item short form, and 0.97 for the four- and eight-item short form, indicating that short forms provide scores very similar to the scores based on the full item bank. For the eight-item short form, reliability is above 0.9 between a *T* score of 22 and 64, and above 0.8 between 20 and 70. In the combined clinical and U.S. general population samples, 91% and 96% of respondents were measured with reliability greater than 0.9 and 0.8, respectively. Reliability for the four-item short form is above 0.9 between a *T* score of 26 and 46, and 55 and 59; and it is above 0.8 between *T* scores of 25 and 65 (see Figure 1). For the four-item form, 42% and 92% of respondents were measured with reliability greater than 0.9 and 0.8, respectively. The raw score is intended to be used for conversion to the *T* score and should not be used for any analyses or other purposes.

Validity. All validity results relate to the *T* score (not the raw summary score). As hypothesized, strong positive correlation was found between the newly developed UWRS and the Connor-Davidson resilience scale scores ($r = .78$). In addition, scores were correlated in the hypothesized direction and magnitude with the Flourishing Scale ($r = .61$), PANAS short form positive affect ($r = .52$), PANAS short form negative affect ($r = -0.47$), PROMIS Depression ($r = -0.56$), and PROMIS physical function ($r = .09$; all correlations are provided in Table 4). These results provide initial evidence supporting validity of the new resilience scale.

Discussion

Resilience has been identified as a potentially important target for therapeutic interventions. The availability of conceptually and psychometrically sound, person-centered, brief and flexible instruments is essential for studying efficacy of treatments aimed at improving resilience, examining the relationships between resil-

Table 3
Summary of University of Washington Resilience Scale (UWRS) T-Score Means and Standard Deviations by Age, Diagnostic Group, and Sex

Variable	Clinical calibration sample		U.S. general population sample	
	n	*M* ± *SD*	*n*	*M* ± *SD*
Overall sample	1,457	50.7 ± 9.9	300	50.0 ± 9.9
Age groups (years)				
18–34	35	53.3 ± 9.2	100	50.2 ± 9.7
35–44	83	49.8 ± 10.7	47	49.0 ± 8.1
45–54	240	49.5 ± 10.5	43	49.5 ± 10.3
55–64	449	50.4 ± 10.0	59	48.6 ± 10.8
65–74	437	51.5 ± 9.3	35	51.5 ± 9.7
75+	212	50.5 ± 9.6	16	54.6 ± 11.0
Disability group				
Spinal cord injury	349	52.0 ± 10.2		N/A
Postpolio syndrome	383	51.5 ± 9.0		
Muscular dystrophy	274	49.1 ± 10.1		
Multiple sclerosis	451	49.9 ± 10.0		
Sex				
Women	946	50.1 ± 9.8	154	49.1 ± 10.5
Men	511	51.7 ± 9.9	146	51.0 ± 9.1

ience and other health domains, and comparing resilience across populations and studies. Such an instrument would also be useful in identifying people at risk for poor outcomes following an adverse event as well as in the context of ongoing stressors.

The results of the present study indicate that the UWRS, whether administered by CAT or short form, is a flexible, conceptually sound measure of resilience, and provides reliable and valid scores. The final item bank includes 28 items calibrated to IRT with the scores on a T-metric ($M = 50$, $SD = 10$). A mean of 50 represents the mean resilience in the U.S. general population. A higher score represents more resilience. Eight- and four-item short forms are available, and their scores are highly correlated with the score based on all items in the item bank ($r \geq .94$). Reliability is excellent (>.90) up to 1.5 *SD* above the mean (the score of 65) when administered by CAT or the eight-item short form. The four-item short form has excellent reliability up to 1 *SD* above the mean (the score of 60). The scores are slightly less reliable for people with high resilience. This is clinically appropriate because higher reliability is much more important for assessing and discriminating resilience levels below the mean in order to identify people may need treatment or other support. The pattern of positive correlations between UWRS scores and measures of resilience, positive affect and physical function, and negative correlations between UWRS scores and depression and negative affect provide strong support for construct validity. Intended for use with any population, the UWRS is also highly relevant for assessing resilience in the context of chronic health conditions and/or disability.

The rigorous development process of the UWRS resulted in several important advancements in measurement of resilience.

First, because we used IRT, the resulting resilience item bank can be administered in a variety of formats, including CAT and short forms, reducing respondent burden while maintaining high reliability.

Second, custom short forms can be developed by selecting different items and different number of items for use in specific research or clinical settings (Cella, Gershon, Lai, & Choi, 2007). For example, for a clinical trial that uses resilience as a primary outcome, researchers may develop a longer short form (e.g., 10 or 12 items) to increase the likelihood that a smaller (but still clinically meaningful) treatment effect could be detected. For different purposes, let's say to use resilience as a covariate in a study aimed at treatment of depression or for population monitoring in an epidemiological study, a four- or five-item short form provides sufficient reliability while keeping the administration time low. Alternatively, if feasible, using CAT would provide an even more precise score at any level of resilience while minimizing respondent burden. Here we want to note that it is not necessary and not recommended to administer all items in the item bank (except perhaps to examine the psychometric properties of the item bank). An important advantage of an IRT-based scale is that regardless of how the UWRS is administered (CAT, eight-/four-item short form or custom short form) and which items are administered, the scores from different versions of the UWRS are directly comparable. However, being comparable does not mean that the scores' reliability is the same. Scores based on short forms with fewer items (e.g., four-item short form) are less reliable, especially at higher levels of resilience than the scores based on short forms with more items (e.g., eight-item short form). For instance, 42% of individuals in our samples were measured with greater than 0.9 reliability with four-item short form compared with 91% with eight-item short form. Scores with reliability of more than 0.9 are desirable for comparison of individual scores (as opposed to group means).

Table 4
Correlations Between University of Washington Resilience Scale and Relevant Validity Measures

Variable	UW Resilience Scale-Full Bank	UW Resilience Scale-Short Form	Connor-Davidson Resilience Scale	PANAS positive affect	PANAS negative affect	Flourishing scale	PROMIS depression
UW Resilience Scale-Full Bank	1						
UW Resilience Scale-Short Form	0.96	1					
Connor-Davidson Resilience Scale	0.78	0.77	1				
PANAS positive affect	0.52	0.50	0.49	1			
PANAS negative affect	−0.47	−0.46	−0.49	−0.23	1		
Flourishing scale	0.61	0.60	0.62	0.51	−0.43	1	
PROMIS depression	−0.56	−0.54	−0.61	−0.36	0.61	−0.61	1
PROMIS physical function	0.09	0.08	0.15	0.06	−0.16	0.17	−0.24

Note. PANAS = the Positive and Negative Affect Schedule; PROMIS = Patient Reported Outcomes Measurement Information System.

With higher score reliability smaller samples sizes are needed to detect a treatment effect (Fries, Krishnan, Rose, Lingala, & Bruce, 2011). At the same time, we recognize that respondent burden is a critical issue in many clinical and research settings. This is the reason we provide detailed information on the reliability of the scores, allowing researchers and clinicians to balance their precision and respondent burden needs.

Third, the UWRS was developed to measure resilience in any population. This facilitates comparisons of resilience across different conditions, populations, or countries.

Fourth, the UWRS's *T* score metric is centered on the U.S. general population mean. A mean of 50 represents the resilience level of the U.S. general population. By centering the scores on a general population matched to the US census, researchers and clinicians can readily interpret scores as below or above the U.S. general population mean and compare resilience in different populations to each other or to the U.S. general population. To users outside of the U.S. it may seem like the UWRS is only intended to be used in the U.S. That is not so. Often scales are centered on the mean of the development sample, a mean point that is more difficult to interpret and generalize than a mean of a representative sample of a large country. Centering the scale on the U.S. general population does not affect the use of the scale in different countries. For instance, a difference of 3 points pre- and postintervention has the same meaning and the same statistical significance (or a lack of significance) regardless of the level of mean resilience in a specific country. To make a comparison with the general population meaningful for non-U.S. users, Canadian researchers could, for example, administer the scale to a sample representative of the Canadian general population and compute the mean score of a representative sample of Canadians. For example, let's say a Canadian sample representative of the Canadian population has a mean resilience score of 52, indicating that the Canadian general population reports slightly more resilience than the U.S. general population. Researchers interested in testing whether the mean resilience score of Canadians with, for instance, chronic pain is statistically significantly different from the Canadian general population score, would test whether the chronic pain sample score is statistically significantly different from 52 (rather than the U.S. mean of 50). Country or condition specific norms can be developed to further aid in score interpretation. It is important to note that cultural appropriateness of the item content and functioning of the items in other English-speaking countries will need to be investigated. Translations should also include evaluation of cultural appropriateness.

Fifth, a lack of statistically significant DIF suggests that men and women, people with different levels of education, people with or without disabilities or chronic conditions, and people from different ethnic and racial backgrounds respond to stressors in similar ways. The results provide evidence that resilience, as defined by this study, is a universally applicable construct. Thus, the UWRS is relevant for use with the general population samples and not just samples of individuals with chronic health conditions.

Sixth, the UWRS can be used in clinical settings to identify patients who could benefit from further evaluation. For instance, a clinic may program a six-item short form to be administered on a tablet and responses are immediately available to clinic staff and/or clinicians. Patients with UWRS scores of 45 of lower (i.e., 0.5 *SD* below the mean) could be referred to the clinic psychologist to evaluate whether the patient could benefit from a treatment or other supports.

The results of this study should be viewed in the context of its limitations. One limitation is the use of a convenience sample of people with chronic medical conditions for the focus groups, cognitive interviews, and the calibration analyses, with only a small subset of medical conditions represented. In addition, no effort was made in this study or in the original longitudinal survey study to make the clinical sample representative of all people with chronic medical conditions. In large survey studies, individuals who volunteer to participate in surveys often have higher education and health literacy, greater interest in promoting wellness, better access to care, and may differ from other people with chronic conditions in important ways. However, the general population sample, whose mean was used to determine the UWRS mean score, was matched on gender, age, race, and ethnicity, addressing some of the concerns of generalizability of the score. All items in the item bank were administered to the disability sample and the scores for the short forms were generated using the IRT parameters for the short form items. It is possible, but not very likely, that if only a short form items were administered the respondents would choose different responses.

It is also important to keep in mind that the construct of resilience is complex and dynamic, and understanding it fully will likely require longitudinal studies with repeated measurement to capture its dynamic nature as well as its relationship, if any, to the type of adversity experienced by the respondent.

Additional limitations include lack of information about test–retest reliability and sensitivity to change. Investigating the UWRS's sensitivity to change will require adequately powered longitudinal prospective studies in contexts where resilience is expected to change. Ideally, studies that provide interventions aimed at increasing resilience. Such studies will provide data that can be used to estimate clinically meaningful (rather than just statistically significant) differences in scores. Until the clinically important differences can be estimated, it is reasonable to interpret the difference of 0.5 *SD* (i.e., 5 points) as clinically meaningful (Farivar, Liu, & Hays, 2004). Finally, the instrument requires approximately sixth grade reading level and may not be well understood by people with low literacy as one in six U.S. adults have literacy skills below a fourth grade level (National Center for Education Statistics, 2018).

Conclusion

The UWRS item bank consists of 28 items that provide an IRT-based *T* score centered on the U.S. general population. The UWRS was developed to be particularly relevant and meaningful for measuring resilience in the context of physical disability, but the scale items were written to reflect resilience in general, and the UWRS can therefore be used in any population. The UWRS provides flexible modes of administration (e.g., computer or paper-and-pencil), including CAT and short forms. The UWRS has excellent psychometric properties and the initial results provide strong support for validity and reliability. The UWRS facilitates comparisons across populations and studies, and is suitable for both clinical practice and research. The item bank and short forms are freely and publicly available for use and can be accessed at http://uwcorr.washington.edu/.

References

Alschuler, K. N., Arewasikporn, A., Nelson, I. K., Molton, I. R., & Ehde, D. M. (2018). Promoting resilience in individuals aging with multiple sclerosis: Results from a pilot randomized controlled trial. *Rehabilitation Psychology, 63,* 338–348. http://dx.doi.org/10.1037/rep0000223

American Psychological Association. (2014). *The road to resilience*. Washington, DC: Author. Retrieved from http://www.apa.org/helpcenter/road-resilience.aspx

Amtmann, D., Cook, K. F., Johnson, K. L., & Cella, D. (2011). The PROMIS initiative: Involvement of rehabilitation stakeholders in development and examples of applications in rehabilitation research. *Archives of Physical Medicine and Rehabilitation, 92,* S12–S19. http://dx.doi.org/10.1016/j.apmr.2011.04.025

Battalio, S. L., Silverman, A. M., Ehde, D. M., Amtmann, D., Edwards, K. A., & Jensen, M. P. (2017). Resilience and function in adults with physical disabilities: An observational study. *Archives of Physical Medicine and Rehabilitation, 98,* 1158–1164. http://dx.doi.org/10.1016/j.apmr.2016.11.012

Bjorner, J. B., Smith, K. J., Stone, C., & Sun, X. (2007). *IRTFIT: A macro for item fit and local dependence tests under IRT models*. Lincoln, RI: QualityMetric Incorporated.

Black, R., & Dorstyn, D. (2015). A biopsychosocial model of resilience for multiple sclerosis. *Journal of Health Psychology, 20,* 1434–1444. http://dx.doi.org/10.1177/1359105313512879

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *American Psychologist, 59,* 20–28. http://dx.doi.org/10.1037/0003-066X.59.1.20

Campbell-Sills, L., & Stein, M. B. (2007). Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress, 20,* 1019–1028. http://dx.doi.org/10.1002/jts.20271

Cardiff. (2015). TeleForm Optical Mark Recognition (OMR) (Version 10.8) [Computer Software]. San Jose, CA: Hewlett Packard.

Cella, D., Gershon, R., Lai, J. S., & Choi, S. (2007). The future of outcomes measurement: Item banking, tailored short-forms, and computerized adaptive assessment. *Quality of Life Research, 16,* 133–141. http://dx.doi.org/10.1007/s11136-007-9204-6

Cella, D., Lai, J. S., Nowinski, C. J., Victorson, D., Peterman, A., Miller, D., . . . Moy, C. (2012). Neuro-QOL: Brief measures of health-related quality of life for clinical research in neurology. *Neurology, 78,* 1860–1867. http://dx.doi.org/10.1212/WNL.0b013e318258f744

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48,* 1–29. http://dx.doi.org/10.18637/jss.v048.i06

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,* 1–30. http://dx.doi.org/10.18637/jss.v039.i08

Crane, P. K., Gibbons, L. E., Ocepek-Welikson, K., Cook, K., Cella, D., Narasimhalu, K., . . . Teresi, J. A. (2007). A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. *Quality of Life Research, 16,* 69–84. http://dx.doi.org/10.1007/s11136-007-9185-5

De Ayala, R. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.

DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care, 45,* S12–S21. http://dx.doi.org/10.1097/01.mlr.0000254567.79743.e2

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D., Oishi, S., & Biswas-Diener, R. (2010). New measures of well-being: Flourishing and positive and negative feelings. *Social Indicators Research, 39,* 247–266. http://dx.doi.org/10.1007/s11205-009-9493-y

Earvolino-Ramirez, M. (2007). Resilience: A concept analysis. *Nursing Forum, 42,* 73–82. http://dx.doi.org/10.1111/j.1744-6198.2007.00070.x

Ehde, D. M., Elzea, J. L., Verrall, A. M., Gibbons, L. E., Smith, A. E., & Amtmann, D. (2015). Efficacy of a telephone-delivered self-management intervention for persons with multiple sclerosis: A randomized controlled trial with a one-year follow-up. *Archives of Physical Medicine & Rehabilitaton, 96,* 1945–1958.

Farivar, S. S., Liu, H., & Hays, R. D. (2004). Half standard deviation estimate of the minimally important difference in HRQOL scores? *Expert Review of Pharmacoeconomics & Outcomes Research, 4,* 515–523. http://dx.doi.org/10.1586/14737167.4.5.515

Flora, D., & Thissen, D. (2002). *User's guide for IRTSCORE: Item response theory score approximation software*. Retrieved from http://www.unc.edu/depts/psychology/dthissen/840F14/IRTScore.pdf

Fries, J. F., Krishnan, E., Rose, M., Lingala, B., & Bruce, B. (2011). Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. *Arthritis Research & Therapy, 13,* R147. http://dx.doi.org/10.1186/ar3461

Gershon, R. C., Lai, J. S., Bode, R., Choi, S., Moy, C., Bleck, T., . . . Cella, D. (2012). Neuro-QOL: Quality of life item banks for adults with neurological disorders: Item development and calibrations based upon clinical and general population testing. *Quality of Life Research, 21,* 475–486. http://dx.doi.org/10.1007/s11136-011-9958-8

Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH toolbox for assessment of neurological and behavioral function. *Neurology, 80,* S2–S6. http://dx.doi.org/10.1212/WNL.0b013e3182872e5f

Hahn, E. A., Cella, D., Chassany, O., Fairclough, D. L., Wong, G. Y., & Hays, R. D. (2007). Precision of health-related quality-of-life data compared with other clinical measures. *Mayo Clinic Proceedings, 82,* 1244–1254. http://dx.doi.org/10.4065/82.10.1244

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42,* 377–381. http://dx.doi.org/10.1016/j.jbi.2008.08.010

Howden, L. M., & Meyer, J. A. (2011, May). Age and sex composition: 2010. *2010 Census Briefs*. Retrieved from https://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf

Hu, L.-T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Humes, K. R., Jones, N. A., & Ramirez, R. R. (2011). *Overview of Race and Hispanic Origin: 2010. 2010 Census Briefs*. Retrieved from https://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf

Johnston, M. C., Porteous, T., Crilly, M. A., Burton, C. D., Elliott, A., Iversen, L., . . . Black, C. (2015). Physical disease and resilient outcomes: A systematic review of resilience definitions and study methods. *Psychosomatics, 56,* 168–180. http://dx.doi.org/10.1016/j.psym.2014.10.005

Kim, D., De Ayala, R. J., Ferdous, A. A., & Nering, M. L. (2011). The comparative performance of conditional independence indices. *Applied Psychological Measurement, 35,* 447–471. http://dx.doi.org/10.1177/0146621611407909

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) for Navy enlisted personnel. *Institute for Simulation and Training, 56*. https://stars.library.ucf.edu/istlibrary/56

Koelmel, E., Hughes, A. J., Alschuler, K. N., & Ehde, D. M. (2017). Resilience mediates the longitudinal relationships between social support and mental health outcomes in multiple sclerosis. *Archives of*

Physical Medicine & Rehabilitation Medicine, 98, 1139–1148. http://dx.doi.org/10.1016/j.apmr.2016.09.127

Leppin, A. L., Gionfriddo, M. R., Sood, A., Montori, V. M., Erwin, P. J., Zeballos-Palacios, C., . . . Tilburt, J. C. (2014). The efficacy of resilience training programs: A systematic review protocol. *Systematic Reviews, 3,* 20. http://dx.doi.org/10.1186/2046-4053-3-20

Loprinzi, C. E., Prasad, K., Schroeder, D. R., & Sood, A. (2011). Stress Management and Resilience Training (SMART) program to decrease stress and enhance resilience among breast cancer survivors: A pilot randomized clinical trial. *Clinical Breast Cancer, 11,* 364–368. http://dx.doi.org/10.1016/j.clbc.2011.06.008

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the Positive and Negative Affect Schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences, 27,* 405–416. http://dx.doi.org/10.1016/S0191-8869(98)00251-7

Manning, L. K., Carr, D. C., & Kail, B. L. (2016). Do higher levels of resilience buffer the deleterious impact of chronic illness on disability in later life? *The Gerontologist, 56,* 514–524. http://dx.doi.org/10.1093/geront/gnu068

Microsoft Corporation. (2016). Microsoft Word (Version 1804) [Computer Software]. Redmond, WA: Microsoft Corporation. Retrieved from https://www.microsoft.com/en-us/

Molton, I. R., Terrill, A. L., Smith, A. E., Yorkston, K. M., Alschuler, K. N., Ehde, D. M., & Jensen, M. P. (2014). Modeling secondary health conditions in adults aging with physical disability. *Journal of Aging and Health, 26,* 335–359. http://dx.doi.org/10.1177/0898264313516166

Moore, R. C., Eyler, L. T., Mausbach, B. T., Zlatar, Z. Z., Thompson, W. K., Peavy, G., . . . Jeste, D. V. (2015). Complex interplay between health and successful aging: Role of perceived stress, resilience, and social support. *The American Journal of Geriatric Psychiatry, 23,* 622–632. http://dx.doi.org/10.1016/j.jagp.2014.08.004

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide. 7th edition*. Los Angeles, CA: Author.

National Center for Education Statistics. (2018). *National Assessment of Adult Literacy: Demographics*. Retrieved from https://nces.ed.gov/naal/kf_demographics.asp

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill, Inc.

Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S - X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement, 27,* 289–298. http://dx.doi.org/10.1177/0146621603027004004

Petriwskyj, A., Parker, D., O'Dwyer, S., Moyle, W., & Nucifora, N. (2016). Interventions to build resilience in family caregivers of people living with dementia: A comprehensive systematic review. *JBI Database of Systematic Reviews and Implementation Reports, 14,* 238–273. http://dx.doi.org/10.11124/JBISRIR-2016-002555

Pilkonis, P. A., Choi, S. W., Reise, S. P., Stover, A. M., Riley, W. T., & Cella, D. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): Depression, anxiety, and anger. *Assessment, 18,* 263–283. http://dx.doi.org/10.1177/1073191111411667

R Core Team. (2016). R: A Language and Environment for Statistical Computing (Version 3.3.1) [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., . . . Cella, D. (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,* S22–S31. http://dx.doi.org/10.1097/01.mlr.0000250483.85507.04

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552–566. http://dx.doi.org/10.1037/0033-2909.114.3.552

Reivich, K. J., Seligman, M. E., & McBride, S. (2011). Master resilience training in the U.S. Army. *American Psychologist, 66,* 25–34. http://dx.doi.org/10.1037/a0021897

Rose, M., Bjorner, J. B., Gandek, B., Bruce, B., Fries, J. F., & Ware, J. E., Jr. (2014). The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *Journal of Clinical Epidemiology, 67,* 516–526. http://dx.doi.org/10.1016/j.jclinepi.2013.10.024

Samejima, F. (1996). The graded response model. In W. J. v. d. Liden & R. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85–100). New York, NY: Springer.

SAS Institute, Inc. (2012–2013). SAS 9.3. Help and Documentation. Cary, NC: SAS Institute Inc.

Schalet, B. D., Hays, R. D., Jensen, S. E., Beaumont, J. L., Fries, J. F., & Cella, D. (2016). Validity of PROMIS physical function measured in diverse clinical samples. *Journal of Clinical Epidemiology, 73,* 112–118. http://dx.doi.org/10.1016/j.jclinepi.2015.08.039

Schalet, B. D., Pilkonis, P. A., Yu, L., Dodds, N., Johnston, K. L., Yount, S., . . . Cella, D. (2016). Clinical validity of PROMIS depression, anxiety, and anger across diverse clinical samples. *Journal of Clinical Epidemiology, 73,* 119–127. http://dx.doi.org/10.1016/j.jclinepi.2015.08.036

Schotanus-Dijkstra, M., Ten Klooster, P. M., Drossaert, C. H., Pieterse, M. E., Bolier, L., Walburg, J. A., & Bohlmeijer, E. T. (2016). Validation of the Flourishing Scale in a sample of people with suboptimal levels of mental well-being. *BMC Psychology, 4,* 12. http://dx.doi.org/10.1186/s40359-016-0116-5

Silverman, A. M., Molton, I. R., Alschuler, K. N., Ehde, D. M., & Jensen, M. P. (2015). Resilience predicts functional outcomes in people aging with disability: A longitudinal investigation. *Archives of Physical Medicine & Rehabilitation Medicine, 96,* 1262–1268. http://dx.doi.org/10.1016/j.apmr.2015.02.023

Silverman, A. M., Verrall, A. M., Alschuler, K. N., Smith, A. E., & Ehde, D. M. (2017). Bouncing back again, and again: A qualitative study of resilience in people with multiple sclerosis. *Disability and Rehabilitation, 39,* 14–22.

Smith, A. E., Molton, I. R., & Jensen, M. P. (2016). Self-reported incidence and age of onset of chronic comorbid medical conditions in adults aging with long-term physical disability. *Disability and Health Journal, 9,* 533–538. http://dx.doi.org/10.1016/j.dhjo.2016.02.002

Southwick, S. M., Bonanno, G. A., Masten, A. S., Panter-Brick, C., & Yehuda, R. (2014). Resilience definitions, theory, and challenges: Interdisciplinary perspectives. *European Journal of Psychotraumatology*. Advance online publication. http://dx.doi.org/10.3402/ejpt.v5.25338

Steptoe, A., Dockray, S., & Wardle, J. (2009). Positive affect and psychobiological processes relevant to health. *Journal of Personality, 77,* 1747–1776. http://dx.doi.org/10.1111/j.1467-6494.2009.00599.x

Stewart, D. E., & Yuen, T. (2011). A systematic review of resilience in the physically ill. *Psychosomatics, 52,* 199–209. http://dx.doi.org/10.1016/j.psym.2011.01.036

Terrill, A. L., MacKenzie, J. J., Einerson, J., & Reblin, M. (2019). Resilience and disability: Intra-personal, interpersonal, and social environment factors. In D. S. Dunn (Ed.), *Disability: Social psychological perspectives* (pp. 313–331). New York, NY: Oxford University Press.

Terrill, A. L., Molton, I. R., Ehde, D. M., Amtmann, D., Bombardier, C. H., Smith, A. E., & Jensen, M. P. (2016). Resilience, age, and perceived symptoms in persons with long-term physical disabilities. *Journal of Health Psychology, 21,* 640–649. http://dx.doi.org/10.1177/1359105314532973

Thissen, D. (2000). Reliability and measurement precision. In H. Wainer & N. J. Dorans (Eds.), *Computerized adaptive testing: A primer* (2nd ed., pp. 159–184). Mahwah, NJ: Erlbaum Publishers.

Toluna. (2017). *PanelBook 2017*. Retrieved from http://www.toluna-group.com/docs/default-source/default-document-library/panelbook_2017_final.pdf?sfvrsn=6

Tulsky, D. S., Kisala, P. A., Victorson, D., Carlozzi, N., Bushnik, T., Sherer, M., . . . Cella, D. (2016). TBI-QOL: Development and calibration of item banks to measure patient reported outcomes following traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 31,* 40–51. http://dx.doi.org/10.1097/HTR.0000000000000131

Tulsky, D. S., Kisala, P. A., Victorson, D., Tate, D. G., Heinemann, A. W., Charlifue, S., . . . Cella, D. (2015). Overview of the Spinal Cord Injury—Quality of Life (SCI-QOL) measurement system. *The Journal of Spinal Cord Medicine, 38,* 257–269. http://dx.doi.org/10.1179/2045772315Y.0000000023

Victorson, D., Tulsky, D. S., Kisala, P. A., Kalpakjian, C. Z., Weiland, B., & Choi, S. W. (2015). Measuring resilience after spinal cord injury: Development, validation and psychometric characteristics of the SCI-QOL Resilience item bank and short form. *The Journal of Spinal Cord Medicine, 38,* 366–376. http://dx.doi.org/10.1179/2045772315Y.0000000016

Willis, G. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412983655

Windle, G. (2010). The Resilience Network: What is resilience? A systematic review and concept analysis. *Reviews in Clinical Gerontology, 21,* 1–18.

Windle, G., Bennett, K. M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes, 9,* 8. http://dx.doi.org/10.1186/1477-7525-9-8

Zautra, A. J. (2009). Resilience: One part recovery, two parts sustainability. *Journal of Personality, 77,* 1935–1943. http://dx.doi.org/10.1111/j.1467-6494.2009.00605.x

Zumbo, B. D. (1999). *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.

Received January 10, 2019
Revision received September 30, 2019
Accepted December 19, 2019 ■