

# National Institutes of Health Toolbox Emotion Battery for English- and Spanish-speaking adults: normative data and factor-based summary scores

Ida Babakhanyan<sup>1,2</sup>  
Benjamin S McKenna<sup>2</sup>  
Kaitlin B Casaletto<sup>3</sup>  
Cindy J Nowinski<sup>4</sup>  
Robert K Heaton<sup>2</sup>

<sup>1</sup>Defense and Veteran's Brain Injury Center, Camp Pendleton, <sup>2</sup>Department of Psychiatry, University of California, San Diego, La Jolla, <sup>3</sup>Department of Neurology, University of California, San Francisco, San Francisco, CA, <sup>4</sup>Department of Medical Social Sciences, Northwestern University, Chicago, IL, USA

**Background:** The National Institutes of Health Toolbox Emotion Battery (NIHTB-EB) is a “common currency”, computerized assessment developed to measure the full spectrum of emotional health. Though comprehensive, the NIHTB-EB's 17 scales may be unwieldy for users aiming to capture more global indices of emotional functioning.

**Methods:** NIHTB-EB was administered to 1,036 English-speaking and 408 Spanish-speaking adults as a part of the NIH Toolbox norming project. We examined the factor structure of the NIHTB-EB in English- and Spanish-speaking adults and developed factor analysis-based summary scores. Census-weighted norms were presented for English speakers, and sample-weighted norms were presented for Spanish speakers.

**Results:** Exploratory factor analysis for both English- and Spanish-speaking cohorts resulted in the same 3-factor solution: 1) negative affect, 2) social satisfaction, and 3) psychological well-being. Confirmatory factor analysis supported similar factor structures for English- and Spanish-speaking cohorts. Model fit indices fell within the acceptable/good range, and our final solution was optimal compared to other solutions.

**Conclusion:** Summary scores based upon the normative samples appear to be psychometrically supported and should be applied to clinical samples to further validate the factor structures and investigate rates of problematic emotions in medical and psychiatric populations.

**Keywords:** emotional functioning, NIH Toolbox Emotion Battery, factor analyses, summary scores, normative data

## Introduction

The National Institutes of Health (NIH) Toolbox Assessment of Neurological and Behavioral Function ([www.nihtoolbox.org](http://www.nihtoolbox.org)) is a set of brief measures that assess cognitive, emotional, motor, and sensory functions across the life span. It was commissioned by the NIH Blueprint for Neuroscience Research to provide a widely accessible, easy to administer, brief method assessing multiple aspects of health in a way that can be uniform across neurological research.<sup>1</sup> Because the battery provides a “common currency” across clinical and research settings, it can be used to monitor neurological and behavioral functioning across the life span with various health conditions and their treatments. The NIH Toolbox Emotion Battery (NIHTB-EB) was created in response to consensus from an expert panel identifying the need to measure both positive and negative aspects of emotions in a standardized manner.<sup>2</sup> NIHTB-EB evolved out of the Patient-Reported Outcomes Measurement Information System (PROMIS). The PROMIS battery focused on the impact of chronic conditions on health-related quality of life (HRQL).<sup>3</sup> At the time, PROMIS included items on depression, anxiety, and

Correspondence: Robert K Heaton  
Department of Psychiatry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA  
Tel +1 858 534 4044  
Fax +1 858 534 9917  
Email [rheaton@ucsd.edu](mailto:rheaton@ucsd.edu)

emotional distress.<sup>4</sup> Recognizing the full spectrum of emotional life and its impact on health, the NIH Toolbox mandate was to develop an assessment tool with a broad focus rather than only assessing negative emotions.

Leveraging the decades of work characterizing the relationship between emotional functioning and health, an expert panel of investigators funded by the NIH identified four theoretically relevant subdomains for inclusion in the NIHTB-EB: negative affect, psychological well-being, stress and self-efficacy, and social relationships.<sup>5</sup> Specifically, given that negative and positive emotions are relatively independent of each other and not necessarily opposite extremes of one continuum,<sup>6–11</sup> the NIHTB-EB aimed to assess negative and positive psychological functioning separately. Additionally, there is a strong, bidirectional relationship between social relationships and emotional health;<sup>12</sup> therefore, the NIHTB-EB aimed to tap into the interpersonal aspects of everyday life, such as support and friendship. Finally, perceptions of stress and self-efficacy significantly impact physical health and mental health both directly (eg, adverse physical effects of stress-related cortisol) and indirectly (eg, selection and application of coping strategies)<sup>13–15</sup> and were therefore considered for inclusion in the final battery.

After the theoretically relevant domains were identified, the committee for the development of the NIHTB-EB was tasked with the selection of psychometrically sound and nonproprietary measures, as well as generation of item banks to measure each of these important constructs when an already existing measure was unavailable. Expert feedback and literature review informed the selection of the item banks for the different scales of the NIHTB-EB.<sup>16</sup> For example, the team of researchers who worked on the negative affect scales included items from the PROMIS item bank and other well-known measures specific to negative emotions.<sup>6</sup> Selections were then made on all of the items that were to be included, and these items went through extensive calibration to promote the Toolbox agenda focused on creating a useful and efficient tool to assess emotions.

Although much thought and consideration went into the selection of the items within each domain, there has not been a comprehensive study within the large normative database, examining the specific domains that the final 17 individual scales represent. There has also been no method proposed for obtaining summary scores for the respective domains. The purpose of this study was to evaluate and compare the factor structure of the NIHTB-EB scales in English- and Spanish-speaking adults through exploratory and confirmatory factor analyses, as well as to begin exploring sociodemographic effects on the battery. Our goal was to identify composite

scales based on the factor analyses findings and provide formulas such that the composite measures may be implemented across research and clinical settings that utilize the NIHTB-EB. Census-weighted normative data were provided for English-speaking adults and sample-weighted norms for Spanish speakers.

## Methods

### Participants and procedures

The NIH Toolbox normative sample of adults consisted of healthy community-dwelling individuals 18–85 years old who were recruited across 10 testing sites using a stratified sampling strategy (strata: age, gender, primary language).<sup>17</sup> Potential study participants were randomly selected from existing databases and completed a telephone screen to determine eligibility based on sociodemographic and linguistic categories. Additional participant inclusion criteria included 1) community-dwelling and noninstitutionalized, 2) ability to follow instructions in English or Spanish, and 3) having adequate physical capability (visual, auditory, vestibular, and motor functions) either independently or with assistive devices, to complete the full Toolbox battery (including also the cognition, motor, and sensory modules).<sup>18</sup> Notably, included adults were presumed to be healthy but who were not explicitly screened or excluded for psychiatric history. Research associates who went through training and certification processes, overseen by a team from Northwestern University, conducted structured interviews to help identify those who could be included in the normative project. Certifiers at Northwestern University had the role of site monitors and supervised all aspects of data collection from set up of data collection to quality assurance throughout the data gathering process. This study complied with the ethical rules for human experimentation stated in the Declaration of Helsinki, with Northwestern University's institutional review board's approval, and written informed consent was obtained from all participants.

Participants included in the NIHTB-EB analysis were 1,036 English-speaking and 408 Spanish-speaking adults who self-identified demographic characteristics (Table 1). All participants who completed the battery in Spanish self-identified their ethnicity as Hispanic. In the English-speaking cohort, 67% identified their ethnicity as non-Hispanic White, 15% as non-Hispanic Black, 13% as Hispanic, and 5% as non-Hispanic others. Demographic comparisons between English and Spanish battery completers revealed that the Spanish sample was younger, with lower education and annual household incomes ( $P$ 's<0.001). Spanish and English speakers were comparable on gender.

A subset of individuals ( $n=235$ , 128 English speakers and 107 Spanish speakers) provided additional demographic information regarding social factors such as marital status, number of children, and social interactions defined by number of people with whom one interacts within a 2-week period. When comparing this subset of individuals with the larger group, those who provided these additional variables were younger (mean = 42.9 years,  $SD=15.4$ ; versus mean = 48.6,  $SD=18.6$ ;  $P<0.001$ ) and were more likely to be female ( $\chi^2[1, N=1,444]=4.6, P=0.03$ ) with fewer years of education (mean = 12.3 years,  $SD=4.2$ ; versus mean = 13.3,  $SD=3.4$ ;  $P<0.001$ ). When comparing English and Spanish speakers on the additional sociodemographic variables, groups did not differ on marital status; however, Spanish speakers had more children ( $P=0.01$ ) and reported having fewer social interactions in a 2-week time period ( $P=0.002$ ).

## Toolbox Emotion Battery

The NIHTB-EB for adults is a computerized assessment of emotions with 17 scales and four theoretically driven subdomains, developed based on psychometric analyses and consistency with the NIH Toolbox purpose (Table 2).<sup>5</sup> The battery takes ~20–30 minutes to complete, and it is self-administered. Detailed descriptions of the NIHTB-EB scales are included in the NIH Toolbox Score and Interpretation

**Table 1** Sample characteristics of total adult sample and subsample with additional sociodemographic data (mean, SD, %)

Sample characteristics	English, n=1,036	Spanish, n=408	P-value
Age	49.11 (18.57)	44.10 (16.72)	<0.0001
Education (years)	14.04 (2.59)	10.74 (4.35)	<0.0001
Gender (% female)	63.22	64.95	0.5382
Household income <\$40k (%)	12.36	44.36	<0.0001
	<b>n=128</b>	<b>n=107</b>	
Marital status (% married)	61.72	67.29	0.3742
Social interaction	10.97 (6.44)	8.67 (4.63)	0.0023
Number of children	1.93 (1.42)	2.43 (1.64)	0.0126

**Notes:** Household income is equivalent to total annual household income. Social interaction is defined by the number of people interacts within a 2-week time frame and includes friends, family, members from church, and coworkers. Total  $N=1,444$ .

**Table 2** NIH Emotion Battery scales and original theoretically identified subdomains

Stress and self-efficacy	Psychological well-being	Negative affect	Social relationships
Perceived stress	Positive affect	Anger physical aggression	Emotional support
Self-efficacy	Life satisfaction	Anger hostility	Instrumental support
	Meaning and purpose	Fear somatic arousal	Friendship
		Sadness	Loneliness
		Apathy	Perceived hostility
		Fear affect	Perceived rejection

**Notes:** Life satisfaction is also called general life satisfaction. Used with permission ©2006–2017 National Institutes of Health and Northwestern University.<sup>35</sup>

**Abbreviation:** NIH, National Institutes of Health.

Guide ([www.nihtoolbox.org](http://www.nihtoolbox.org)) and are summarized in Table 3 for the individual scales as well as for the final, confirmatory factor analyses (CFA)-based summary scores. Each item administered has a 5- or 7-point Likert scale with options ranging from “not at all” to “very much”. Each scale is scored using item response theory (IRT) methods, producing an IRT generated theta score. In IRT, the assumption is that all individuals have some degree of the underlying trait and the amount of that trait determines the probability that they will answer an item in a specific way.<sup>19</sup> Additionally, the battery is computer adaptive to accurately and efficiently assess each latent construct. This means that the items that an individual participant receives are dependent on his/her prior responses and therefore highly individualized to sensitively capture his/her emotional functioning; due to this approach, not all participants complete the exact same set of individual items. Scores more than one standard deviation below the mean ( $T<40$ ) suggest low level of the trait, and scores of more than one standard deviation above the mean ( $T>60$ ) suggest high level of the trait. All scales of the NIHTB-EB are freely available on the HealthMeasures.net website, and the correct page can be directly accessed using <http://www.healthmeasures.net/explore-measurement-systems/nih-toolbox/obtain-and-administer-measures>. Under “Obtain and Administer Measures”, select “Download a zip file of all available NIH Toolbox Emotion PDFs” and then open the zip file, open the “English” file, and select “Self-Report 18+” to view all scales in the adult battery.

## Derivation of 2010 US census-weighted normalized T-scores

We determined that normative adjustments for age and other demographic variables were not theoretically desirable or statistically necessary for the NIHTB-EB scores. That is, emotion scores are most usefully interpreted as reflecting the absolute amount of the trait in an individual, not the relative amount of the trait compared to others of that individual’s age or gender; additionally, we found that demographics were minimally associated with the NIHTB-EB scores (ie, <5%

**Table 3** Summary descriptions of the NIHTB-EB composites and component scales

Scale/composite	Emotions assessed	Potentially problematic
Psychological well-being (three scales)	Overall positive emotions, feeling content with respect to self and life	$T < 40$
Positive affect	Feelings of contentment, happiness, enthusiasm, joy, and excitement	$T < 40$
General life satisfaction	Feelings of satisfaction with one's self and other aspects of life	$T < 40$
Meaning and purpose	The extent to which individuals feel that their lives are meaningful and/or make sense	$T < 40$
Negative affect (five scales)	Overall negative emotions ranging from anger, fear, depression, and stress	$T > 60$
Anger – affect	Feelings of cynicism and frustration	$T > 60$
Anger – hostility	Feelings of antagonism or animosity	$T > 60$
Fear – affect	Anxiety over perceptions of threat	$T > 60$
Sadness	Feelings of depression	$T > 60$
Perceived stress	A person's view of life challenges in relation to their values and coping resources	$T > 60$
Social satisfaction (five scales)	Social support and perceptions of social relationships	$T < 40$
Perceived rejection	Perceptions of being ignored or others not listening or responding to requests for help	$T > 60^a$
Friendship	Perceptions about the availability of companions or friends with whom to affiliate	$T < 40$
Loneliness	Feelings of being alone or socially isolated from other individuals	$T > 60^a$
Emotional support	The perception that others are available to listen to one's concerns with understanding and caring	$T < 40$
Instrumental support	The perception that people are available to provide information or advice needed to overcome problems	$T < 40$
Supplemental scales		
Anger – physical aggression	Aggressive behaviors	$T > 60$
Fear – somatic arousal	Anxiety with autonomic arousal	$T > 60$
Perceived hostility	Perceptions of being criticized or berated	$T > 60$
Self-efficacy	Feelings of adequate control over one's life, and confidence in being able to manage one's own functioning and outcomes	$T < 40$

**Notes:** <sup>a</sup>Reverse coded for summary score computation.  $T=T$  score. Used with permission ©2006-2017 National Institutes of Health and Northwestern University.<sup>35</sup>

**Abbreviation:** NIHTB-EB, National Institutes of Health Toolbox Emotion Battery.

variance was accounted for on each scale). However, in order to ensure that the normative sample was as representative of the general US population as possible, we weighted our sample to reflect the demographics of the 2010 US census for English speakers. To achieve this, we applied raking procedure<sup>20</sup> using Statistical Analysis System macro “raking” by Battaglia et al.<sup>21</sup> This method assigns a weight, which is demographically proportionate to US 2010 census data, based on a participant's age, gender, education, and race/ethnicity.

For individual scales, raw (theta) scores for each scale in the census-weighted sample were converted to sample-based normalized  $T$ -scores ( $T=50$ ;  $SD=10$ ). Therefore, the normalized  $T$ -scores represent an individual's emotional characteristics compared to the average English-speaking person in the USA.<sup>18</sup> For the Spanish-speaking cohort, raw (theta) scores were converted to sample-based  $T$ -scores without census-weighted corrections, given that there was no appropriate census data for this cohort. Therefore, normalized  $T$ -scores on the Spanish NIHTB-EB represent an individual's affective characteristics compared to our large normative cohort of Spanish-speaking adults.<sup>22</sup>

## NIHTB-EB factor analyses

In order to create summary scores that reflect the underlying latent structure of the NIHTB-EB, factor analyses were conducted using single sample cross-validation methodologies. Specifically, English- and Spanish-speaking samples were split into two samples within each group stratified on gender and age. For English speakers, one subsample ( $n=636$ ) was used for exploratory factor analyses (EFA) and another subsample ( $n=400$ ) was used for CFA. Similarly, for Spanish speakers one sub-sample ( $n=208$ ) was used for EFA and the other subsample ( $n=200$ ) for CFA. In this way, the latent constructs underlying the NIHTB-EB scales could be examined with EFA and validated with CFA in a separate sample. All factor analyses were performed on raw (theta) scores for English- and Spanish-speaking cohorts separately, using the R software and the “lavaan” package.<sup>23</sup>

To identify underlying latent factors, EFA with maximum likelihood estimation was used to calculate eigenvalues and determine the number of factors to extract using multiple approaches. A multiple approach to data reduction, rather than use of single criteria (eg, scree test, eigenvalues  $>1$ , and

cumulative percent of variance extracted), has been suggested to be the best practice in EFA research.<sup>24</sup> The eigenvalues were obtained from a principal components analysis. A conservative approach was initially taken for factor extraction. If fewer than the appropriate number of factors are initially extracted, the factors may include excessive errors due to important variables going unnoticed. The salient loading criterion adds 50% to what is suggested by eigenvalue criteria. Strength of the scale loadings on each factor was examined, and factors with a minimum of three scales loading  $>0.3$  or 2 scales loading  $>0.5$  were retained. Also, consistent with salient loading criterion, scales that did not demonstrate a minimum of a .13 margin from the factor it loaded the highest on were removed for the analysis until there were no cross-loadings within a .13 margin. An oblique promax rotation of the extracted factors was utilized to achieve the simplest structure. Inter-item correlations and Cronbach's  $\alpha$  were examined to calculate internal consistency estimates of reliability. Seventeen scales were entered into the EFA.

To validate the best-fitting models determined from a priori hypotheses and the EFA step, CFAs were performed. Specifically, the latent structure of the theoretically pre-existing subdomains (4-factor solution), a 1-factor (all scales), 2-factor (positive and negative scales), and the factor solution derived from the EFA step were examined with a CFA approach (refer Table S1 for the specific scales within each factor solution). The distributions for each of the 17 scales were first examined for normality. CFA for each factor model was conducted using maximum likelihood estimation with robust (Huber-White) standard errors while also modeling correlation among factors. Use of the chi-square likelihood ratio test to assess model fit has been deemed unsatisfactory for numerous reasons.<sup>25</sup> Rather, many researchers have suggested the use of multiple measures of model fit.<sup>26</sup> Therefore, the following measures of model fit were used: 1) the comparative fit index (CFI),<sup>27</sup> which compares the target model to a baseline null model that specifies no factors (values  $>0.90$  indicate adequate model fit and values  $>0.93$  indicate good model fit); 2) the root mean square error of approximation (RMSEA),<sup>28</sup> which adjusts fit by weighting values by the number of parameters estimated (values  $<0.08$  indicate adequate model fit while  $<0.05$  indicate good model fit);<sup>29</sup> and 3) standardized root mean square residual (SRMR),<sup>30</sup> which is an absolute measure of fit defined as the standardized difference between the observed correlation and the predicted correlation (values  $<0.08$  indicate good model fit). Using these indices, the best fitting and most parsimonious factor model were identified.

To maximize model fit, we revised the best fitting model using the Wald test,<sup>31</sup> which identifies scales that if dropped would improve overall model fit, and proceeded to examination of the standardized factor loadings for each scale.

## Summary score creation

We used the best fitting model from CFA to create summary scores in the full sample, which included all participants. The full sample was used in this step to provide the most precise estimates in our summary score equations ( $N=1,026$  for English and  $N=408$  for Spanish). Specifically, summary scores were created by weighting the raw (theta) score for each participant's individual scale by the CFA standardized factor loadings and then averaging across scales within a latent domain. The weighted average scores were then normalized to a  $T$ -score distribution (mean 50 and standard deviation 10) similar to how individual normalized scales were created.

## Potentially problematic emotion cut-point

We established cut-points of more than one standard deviation below the mean ( $T < 40$ ) for positive emotion scales and more than one standard deviation above the mean ( $T > 60$ ) for negative emotion scales to indicate a "potentially problematic" emotion across the summary scores (refer Table 3 for each scale's problematic direction).<sup>32</sup> Using the normal curve, we expect such a cut-point to demonstrate ~84% specificity (ie, ~16% potentially problematic emotion) among a general population of healthy individuals.

To help control for Type I error due to large sample sizes and multiple analyses, a somewhat conservative  $\alpha$  value of 0.01 was used to indicate significance for all analyses.

## Results

### Exploratory factor analyses

The EFA of the 17 scales for the stratified sample ( $n=636$  English speakers,  $n=208$  Spanish speakers) supported the same 3-factor solution for the English- and Spanish-speaking cohorts. Seven scales (fear affect, anger affect, sadness, perceived stress, anger hostility, fear somatic arousal, and anger physical aggression) loaded saliently on Factor 1 (negative affect). Five scales (friendship, emotional support, instrumental support, and reverse-scored loneliness and perceived rejection) loaded saliently on Factor 2 (social satisfaction). Three scales (meaning, life satisfaction, and positive affect) loaded saliently on Factor 3 (psychological well-being). Self-efficacy and perceived hostility did not load saliently for either cohort (Table 4).



**Table 4** Oblique rotated factor loadings of exploratory factor analysis from split sample

Emotion Battery Scales	English (N=636)			Spanish (N=208)		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
Fear affect	<b>0.81</b>	0	0	<b>0.88</b>	0.04	-0.01
Anger affect	<b>0.81</b>	0.05	-0.04	<b>0.81</b>	0.05	-0.08
Sadness	<b>0.70</b>	-0.06	-0.13	<b>0.79</b>	-0.15	-0.01
Perceived stress	<b>0.72</b>	0.03	-0.24	<b>0.60</b>	-0.03	-0.26
Anger hostility	<b>0.47</b>	-0.03	-0.26	<b>0.55</b>	-0.05	-0.25
Fear somatic arousal	<b>0.46</b>	-0.07	0.02	<b>0.64</b>	0.07	0.01
Anger physical aggression	<b>0.41</b>	-0.05	0.04	<b>0.43</b>	-0.12	-0.16
Perceived hostility	0.33	<b>-0.49</b>	0.16	-0.42	<b>-0.40</b>	-0.03
Loneliness	0.24	<b>-0.57</b>	0.10	-0.29	<b>-0.61</b>	0.03
Perceived rejection	0.25	<b>-0.72</b>	0.16	-0.24	<b>-0.79</b>	-0.13
Friendship	0.12	<b>0.66</b>	0.21	0.12	<b>0.57</b>	0.20
Emotional support	0.15	<b>0.81</b>	0.13	0.16	<b>0.72</b>	0.27
Instrumental support	0.01	<b>0.61</b>	0.03	0.07	<b>0.63</b>	0.16
Meaning	-0.03	0.07	<b>0.77</b>	0.01	0.07	<b>0.85</b>
Life satisfaction	-0.16	0.07	<b>0.66</b>	-0.09	-0.05	<b>0.83</b>
Positive affect	-0.33	0.16	<b>0.51</b>	-0.22	0.21	<b>0.57</b>
Self-efficacy	-0.29	0.13	<b>0.57</b>	-0.13	0.20	<b>0.32</b>

**Note:** Factor loadings in bold designate the factor in which the individual scales are components.

For the English-speaking cohort, Factor 1 explained 23% of the variance (Cronbach's  $\alpha=0.86$ ), Factor 2 explained 18% of the variance (Cronbach's  $\alpha=0.84$ ), and Factor 3 explained 13% of the variance (Cronbach's  $\alpha=0.84$ ). Together, the factor structure accounted for 54% of the total variance. For the Spanish-speaking cohort, Factor 1 explained 24% of the variance (Cronbach's  $\alpha=0.86$ ), Factor 2 explained 17% of the variance (Cronbach's  $\alpha=0.85$ ), and Factor 3 explained 15% of the variance (Cronbach's  $\alpha=0.82$ ). Together, the factor structure accounted for 57% of the total variance.

## Confirmatory factor analyses

As with EFA, the distributions of the 17 scales were adequate for CFA analyses. Table 5 reports the  $\chi^2$  test statistics and model fit indices for all CFA models for English ( $n=400$ ) and Spanish ( $n=200$ ) administered scales. Anger physical aggression and fear somatic arousal were the lowest weighting scales on negative affect (loading  $\sim 0.40$ ), and given that these scales did not improve fit indices, reduced parsimony, and were theoretically peripheral, they were excluded from the final CFA models.

The revised 3-factor model derived from the EFA step was the most parsimonious and best fitting model, as indicated by the CFI, RMSEA, and SRMR indices. Table 6 presents the standardized factor loadings for the best fitting 3-factor model, all of which were significant at  $P<0.001$ . Table 7 presents the correlation matrix among latent variables. For each language sample, negative affect was negatively associated with social satisfaction and psychological well-being.

**Table 5** CFA model fit indices from split samples

CFA model	$\chi^2$ (df)	CFI	RMSEA	SRMR
English (N=400)				
1-Factor model	716.30 (90)	0.79	0.132	0.082
2-Factor model	432.25 (89)	0.89	0.099	0.067
3-Factor model	355.93 (87)	0.91	0.088	0.064
4-Factor model	337.54 (85)	0.91	0.087	0.063
Revised 3-factor model	295.97 (62)	0.92	0.090	0.060
Spanish (N=200)				
1-Factor model	451.46 (90)	0.73	0.144	0.096
2-Factor model	275.26 (89)	0.86	0.104	0.076
3-Factor model	221.72 (87)	0.90	0.089	0.069
4-Factor model	211.48 (85)	0.91	0.088	0.067
Revised 3-factor model	167.84 (62)	0.92	0.086	0.065

**Notes:**  $\chi^2$ , model unscaled chi-square statistic; df, model degrees of freedom.

**Abbreviations:** CFA, confirmatory factor analysis; CFI, comparative fit index; RMSEA, root mean square of approximation; SRMR, standardized root mean square residual.

Also, social satisfaction and psychological well-being were positively associated with each other (Table 7).

To better understand if there were gender differences in our CFA model, we examined gender invariance for each language group separately. Among English speakers, there were no statistical differences between males ( $\chi^2=334.58$ , CFI =0.90, RMSEA =0.102, SRMR =0.061) and females ( $\chi^2=412.71$ , CFI =0.92, RMSEA =0.093, SRMR =0.054) from the results of a  $\chi^2$  test comparing the models ( $P>0.05$ ). Similarly, among Spanish speakers, there were no statistical differences between males ( $\chi^2=445.90$ , CFI =0.90, RMSEA =0.105, SRMR =0.063) and females ( $\chi^2=498.51$ , CFI =0.93, RMSEA =0.088, SRMR =0.050) from the results of a  $\chi^2$

**Table 6** CFA model factor loadings for split sample

Scale	Standardized loadings	
	English	Spanish
Negative affect		
Anger affect	0.797	0.684
Anger hostility	0.646	0.535
Sadness	0.853	0.854
Fear affect	0.838	0.894
Perceived stress	0.802	0.782
Social satisfaction		
Friendship	0.691	0.637
Loneliness	-0.783	-0.776
Emotional support	0.785	0.736
Instrumental support	0.552	0.629
Perceived rejection	-0.658	-0.756
Psychological well-being		
Life satisfaction	0.716	0.785
Meaning	0.740	0.753
Positive affect	0.858	0.898

**Abbreviation:** CFA, confirmatory factor analysis.

**Table 7** CFA model latent variable correlations from split sample

Language group factors	Negative affect	Social satisfaction	Psychological well-being
English			
Negative affect			
Social satisfaction	-0.708		
Psychological well-being	-0.733	0.865	
Spanish			
Negative affect			
Social satisfaction	-0.698		
Psychological well-being	-0.577	0.762	

**Abbreviation:** CFA, confirmatory factor analysis.

test comparing the models ( $P>0.05$ ). However, among both language groups, our revised 3-factor CFA model fit was slightly better for females compared to males.

In an effort to remain consistent with the original work of NIHTB researchers, examination of the scales which make-up each factor and the underlying construct led us to title Factor 1 as negative affect (NA). The scales in Factor 1 have the common theme of negative emotions (fear, anger, sadness, stress). Factor 2 is titled social satisfaction (SS), which included the common theme of the sense of support by others, connection to others, and how one feels others' view him/her. Finally, Factor 3 is called psychological well-being (PWB) with scales that target positive emotions and the common theme of feeling content with aspects of self and life. Table 3 provides brief descriptions of emotions assessed by all individual NIHTB-EB scales and factor-based summary scores.

To test for measurement invariance between English- and Spanish-speaking samples in the 3-factor model, we

examined increasingly restricted models between groups including 1) configural invariance (identical factor structures), 2) weak invariance (factor loadings are constrained to be equal), and 3) strong invariance (factor loadings and intercepts constrained to be equal). Comparisons of models revealed nonsignificant changes in  $\chi^2$  comparing configural to weak invariance ( $\Delta\chi^2=9.59$ ,  $df=10$ ,  $P=0.477$ ) and comparing strong to weak invariance ( $\Delta\chi^2=8.90$ ,  $df=10$ ,  $P=0.542$ ). Additionally, there were no changes in CFI indices in these comparisons but there were small changes in RMSEA ( $\Delta RMSEA=0.004$  for both comparisons). Thus, these findings suggest that the 3-factor model is equivalent between English and Spanish groups.

Finally, we applied the best fitting 3-factor model to the full sample in each language group in order to have the most precise estimates for the purpose of creating summary scores. Parameter estimates from these CFA models were used to generate the summary score equations (Table S2).

## Conversion to *T*-scores

Based on the normative data, Tables S3 and S4 present mean and standard deviation of individual scales, along with formulas for conversion of raw scores to normalized *T*-scores, separately for English and Spanish versions of the of the NIHTB-EB. Values provided are census weighted for English speakers and sample weighted for Spanish speakers.

To determine whether demographic characteristics were significantly associated with results on the NIHTB-EB, the effect of age, education, gender, ethnicity, and household income was evaluated through individual regression analyses for each individual scale, separately for English and Spanish speakers. Significant effect sizes, measured in individual adjusted *R*-squared value, ranged from 0.005 to 0.048 for English speakers and 0.017 to 0.033 for Spanish speakers. Because the results indicated relatively small effect sizes for demographic variables and also because our goal was to provide scores for emotional functioning, which address the question of whether an individual is reporting high or low levels of the specific emotion, we elected not to recommend or provide demographic corrections for the Emotion Toolbox.

Although we are not correcting for age, gender, or education, to an extent we are accounting for the linguistic and associated cultural background influences that may be observed on test performances by providing separate normative formulas for those administered the battery in Spanish and English. Our group will provide details of demographic effects on NIHTB-EB scores, separately by linguistic groups, in a manuscript following this project.

## Summary scores and base rates

Summary scores based on CFA results and factor weights for Spanish and English versions of the battery are provided in Table S2. To establish base rates for potentially problematic emotional functioning, emotional distress was defined by more than one standard deviation beyond the mean in the problematic direction for each scale and composite. Base rates for problematic emotions in the normative sample for the English-speaking cohort revealed 13.9% problematic emotions for negative affect, 16.8% for social satisfaction, and 15.2% for psychological well-being. Base rates of problematic emotions for the Spanish-speaking cohort revealed 18.4% with distress for negative affect, 18.2% for social support, and 13.0% for psychological well-being.

## Social factors and summary scores

A majority of individuals in the combined language samples ( $n=1,083$ ) provided information regarding total annual household income. Individuals with a household income  $\geq$ US\$40,000 reported significantly more social satisfaction (mean = 50.89, SD =9.54; versus mean = 47.97, SD =10.72;  $P<0.001$ ) and psychological well-being (mean = 50.36, SD =9.47; versus mean = 47.65, SD =10.47;  $P=0.0016$ ), as well as slightly less negative affect (mean = 49.42, SD =9.32; versus mean = 51.21, SD =10.93;  $P=0.0169$ ). There was an interaction between income and language for negative affect,  $P=0.030$ . English speakers who had an annual household income  $\geq$ \$40,000 reported significantly less negative affect compared with those with less income (mean = 49.60, SD =9.34; versus mean = 53.51, SD =11.93;  $P<0.001$ ). There was no effect of income on negative affect for Spanish speakers (income  $>$ \$40,000, mean =48.49, SD =9.16; versus income  $<$ \$40,000, mean =49.16, SD =9.57).

A much smaller subset of individuals ( $n=235$ ) provided information on additional sociodemographic variables. Although the psychological well-being summary score was not computed for this subsample due to some missing data on the three scales that make up this factor, summary scores for negative affect and social satisfaction were computed and are available. Relevant to the representativeness of this subsample, their mean negative affect and social satisfaction scores were quite similar to the average results for the total sample (negative affect mean =49.83, SD =9.69; social satisfaction mean =50.22, SD =10.04). Results indicate that individuals who were married reported significantly less negative affect (mean = 48.46, SD =8.63; versus mean = 52.27, SD =10.96;  $P=0.004$ ;  $d=0.39$ ), as well as more social satisfaction (mean = 51.85, SD =9.28; versus mean = 47.32, SD =10.71;  $P<0.001$ ;

$d=0.45$ ) compared with those not married. There was a borderline interaction between marital status and language for social satisfaction,  $P=0.0456$ . For English speakers ( $n=127$ ), being married was associated with greater social satisfaction (mean = 52.00, SD =9.43; versus mean = 45.12, SD =10.39;  $P<0.001$ ;  $d=0.69$ ). However, for Spanish speakers, this was not the case (married, mean =51.68, SD =9.17 versus not married, mean=50.45, SD =10.55;  $d=0.12$ ). Having children was not significantly associated with the two summary scores. The number of individuals with whom one interacts within a 2-week time span also was not significantly associated with negative affect; however, individuals with greater numbers of social interactions reported significantly greater social satisfaction ( $F[1, 228]=16.24$ ,  $P<0.001$ ).

## Discussion

The NIHTB-EB provides a computerized method of briefly assessing a broad spectrum of emotional functioning by including both positive and negative aspects of emotions. Domains were selected by experts and item banks created from the PROMIS battery, already existing well-established nonproprietary measures, as well as new items created where prior measures could not be identified. In the end, 17 scales were developed as the core measures within the adult battery. In this study, we evaluated the domain structure of the NIHTB-EB for both English and Spanish speakers in a project aimed at creating summary scores, which has not been done previously. Here, we present census-weighted norms for the NIHTB-EB English speakers and sample-weighted norms for Spanish speakers. We have provided formulas that can be used to convert raw scores (theta scores provided by the NIHTB Assessment Center program) to standard  $T$ -scores for English and Spanish speakers separately, based on data from the normative samples. Demographically uncorrected scores are provided and, for English speakers, can be interpreted as reflecting an individual's absolute level of that emotion compared to the average English-speaking US adult. These scores are also on a common metric, which may facilitate profile analyses and longitudinal comparisons. We identified three distinct constructs (negative affect, social satisfaction, and psychological well-being) and provided formulas using factor weights from CFA results for computing the summary scores. The final model and summary scores are applicable to both English- and Spanish-speaking adults. Given that we based our corrections on the normal curve, which estimates ~16% of the population will fall one standard deviation above and below the mean, respectively, the base rates on our normative samples are commensurate with expectations



for a normal distribution (refer Table 3 for scale descriptions and domain specifications). Base rates set in the normative sample with these summary scores can be applied to clinical samples to help differentiate problematic emotions across the identified domains.

In the absence of any gold standard assessments for validating cut-points in the current study, we tentatively use the term potentially problematic (not “abnormal”) to interpret scores beyond the one standard deviation point in the direction of distress. Of course, clinicians and researchers are free to set their own cut-points, especially as may be informed by future investigations of the NIHTB-EB. In this regard, however, we would advance the following considerations for NIHTB-EB users: in view of the fact that this battery aims to assess both positive and negative emotions and is intended for use with the general population as well as with clinical samples, it may be too restrictive to use cut-points that would classify almost all nonclinical (or undiagnosed) individuals in the general population as having nonproblematic emotional functioning.

The summary scores presented here can be used across research and clinical settings to aid in more efficient or parsimonious interpretation of findings from NIHTB-EB’s 17 scales. Although greater breadth of information is provided by consideration of all the individual scales, summary/composite scores integrate a significant amount of information into one score and may show greater reliability than the individual component scales. The data points or scales within each composite are now statistically and conceptually related based on the analyses we conducted, and the single score reduced the potential for “information overload”, making the battery more user-friendly. In many situations, a more efficient and user-friendly approach is consistent with the NIH Toolbox objective. Additionally, we have begun to validate these summary scores by demonstrating their association with social variables. For example, greater number of social interactions is associated with an increased sense of social satisfaction on our social satisfaction summary score. The NIH Toolbox initiative has now incorporated these presented normative standards and computed summary scores into the NIHTB-EB iPad scoring program.

It is beyond the scope of the current project to fully address the validity of the NIHTB-EB, other than to report several relevant sociodemographic associations within the NIHTB national norming study. However, validation work with clinical samples is under way, and the findings and norms presented here are intended to be foundational for

such efforts. Given the battery was originally developed in putatively healthy individuals’ representative of the national census, clinical studies across more severe and diverse psychopathologies will importantly inform the criterion validity of the battery. One major strength of the NIHTB-EB is its comprehensive approach to mental health status, including measures across positive and negative affect and social functioning, which may increase its ability to capture and characterize even nuanced differences in psychological functioning across neuropsychiatric disease. Approximately 50 ongoing or completed studies (with >4,400 participants) are registered with the web-based NIHTB Assessment Center and include measures for the Emotion Battery summary scores, and some of these studies are beginning to report results with neurological samples (e.g., spinal cord injury, traumatic brain injury, and stroke).<sup>33</sup> The latter research found significant elevations in negative affect and lower levels of social satisfaction and psychological well-being in individuals with these neurological conditions compared to healthy adults, but also some differences across the neurological conditions. Additionally, the battery demonstrated sensitivity to improvement with treatment (transcranial magnet stimulation) in a recent case study of traumatic brain injury.<sup>34</sup> Nonetheless, given its novelty, continued work to support the sensitivity of the NIHTB-EB to mental health disease is needed. In addition, calculations for the current NIHTB-EB summary scores and norms have been programmed into a recent update of the NIHTB iPad app for use in ongoing and new studies.

There are several limitations in these newly developed normative standards. First, given this normative data is based on the US population and subtle cultural variations have been shown to impact how individuals report emotional health, generalizations cannot be made for international studies at this time. Also, we are not recommending demographic corrections for the Emotion Battery based on relatively small demographic effect sizes and interest for the investigation of emotions as the absolute level of that particular emotion compared to the average person residing in the USA. Interpreting scores of emotional functioning differs from interpretations used for cognitive functioning, which within the neuropsychological context aims to estimate the types and amounts of change in cognition that may have resulted from injury or disease affecting the central nervous system. Accurate classification of neuropsychological impairment, for example, is dependent on the normative comparison applied, such as what is the expected level of cognitive performance if the individual had a healthy brain and never

acquired any central nervous system compromise.<sup>18</sup> Although CNS dysfunction may affect emotional functioning as well, premorbid emotional status (as reflected in the Toolbox normative samples) is much less associated with demographics than is cognition. Nevertheless, we recognize that the current norming process did not take into account subtle effects of demographic variables. We did observe a trend for older individuals to report less negative emotions, for example. These trends can be further explored within specific populations to better understand their stability and significance. Furthermore, in creating summary scores, the RMSEA fit indexes for our final CFA models were not  $<0.05$ , which has been suggested as cutoff for good model fit.<sup>29</sup> However, our other fit indices (ie, CFI and SRMR) suggested that our final models adequately fit the data and produced valid summary scores of emotion in our sample. Nonetheless, future research creating more complex factorial models may yield a more accurate understanding of the underlying latent structure of the Toolbox emotional battery.

In addition, although we were able to separate English and Spanish speakers and provide normative data for each cohort, in the Spanish-speaking cohort, there is variability that could be important to emotional functioning that was not accounted for by the norming project. Information such as country of origin and years since immigration to the USA within the Spanish-speaking cohort was not accounted for. With a larger sample of Spanish speakers, and more comprehensive data collection process that includes items specific to diversity, these factors could be further explored. Also, other potentially important background factors were not consistently assessed in the normative study. Variables specific to social support were not systematically assessed in the norming project, such as socioeconomic and marital status. For example, marital status was available for only ~17% of the sample and was found to be the largest contributor to the emotion scales at the group level (married individuals, as a group, tended to evidence somewhat better emotional health). We plan to report details of (relatively modest) associations with demographic factors in future report.

Also moving forward, application of these normative standards and summary scores with the NIHTB-EB among various clinical populations is warranted to provide validation of the factor structures. A major limitation to this study is the lack of concurrent or discriminant validity for the newly created summary scores. Within the normative sample, we did not have data available to compare the current summary scores with other more established emotional/psychological measures. We are in the process of assessing and reporting

effects of various neurological and psychiatric conditions on the NIHTB-EB, in some cases in relation to the other Toolbox domain instruments (cognition, motor, sensory), and in some cases in relation to other emotion assessments and standardized assessments of current and lifetime histories of various psychiatric conditions (major depressive disorder, substance use disorders, ADHD, and ASPD). However, these projects will have smaller samples and different goals; therefore, they are not within the scope of this study.

Furthermore, research with clinical samples should consider profiles of the NIHTB-EB scores both across and within composite categories. For example, would individuals diagnosed with major depressive disorder (MDD) tend to score in the problematic direction on all three summary scores and, within the negative affect category, will sadness typically be identified as the most problematic? For individuals who are successfully treated for MDD, what patterns of changes will be observed on the NIHTB-EB? Answering similar questions will help validate the NIHTB-EB and the newly constructed scales' construct validity. Solidified construct validity of the measure will increase its utility in clinical settings. This is particularly important given that there are not many methods of assessment for emotions that have a similar broad focus. Summary scores based on the normative samples appear to be psychometrically sound and should be applied to clinical samples to validate the factor structures as well as to investigate rates of problematic emotions in medical and psychiatric populations.

## Acknowledgments

This study was supported by a cooperative agreement from the National Institutes of Health to Northwestern University (U2CCA186878; PI: David Cella, PhD). These contents do not necessarily represent an endorsement by the US Federal Government (refer [www.healthmeasures.net](http://www.healthmeasures.net) for additional information). Funding for HealthMeasures was provided by the National Institutes of Health grant U2C CA186878. We wish to thank Michael Thomas, PhD, for his invaluable consultation on statistical methodologies used in this manuscript.

## Disclosure

The authors report no conflicts of interest in this work.

## References

1. Gershon RC, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. *J Appl Meas*. 2010;11(3):304–314.
2. Nowinski CJ, Victorson D, Debb SM, Gershon RC. Input on NIH toolbox inclusion criteria: surveying the end-user community. *Neurology*. 2013;80(11 suppl 3):S7–S12.

3. Rothrock NE, Hays RD, Spritzer K, Yount SE, Riley W, Cella D. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2010;63(11):1195–1204.
4. Revicki DA, Cook KF, Amtmann D, Harnam N, Chen W-H, Keefe FJ. Exploratory and confirmatory factor analysis of the PROMIS pain quality item bank. *Qual Life Res*. 2014;23(1):245–255.
5. Salsman JM, Butt Z, Pilkonis PA, et al. Emotion assessment using the NIH toolbox. [Miscellaneous Article]. *Neurology*. 2013;80(11 suppl 3):S76–S86.
6. Pilkonis PA, Choi SW, Salsman JM, et al. Assessment of self-reported negative affect in the NIH toolbox. *Psychiatry Res*. 2013;206(1):88–97.
7. Watson D, Tellegen A. Toward a consensual structure of mood. *Psychol Bull*. 1985;98(2):219–235.
8. Diener E, Suh EM, Lucas RE, Smith HL. Subjective well-being: three decades of progress. *Psychol Bull*. 1999;125(2):276–302.
9. Neubauer AB, Voss A. Validation and revision of a German version of the balanced measure of psychological needs scale. *J Individ Differ*. 2016;37(1):56–72.
10. Ryff CD. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *J Pers Soc Psychol*. 1989;57(6):1069–1081.
11. Sheldon KM, Hilpert JC. The balanced measure of psychological needs (BMPN) scale: an alternative domain general measure of need satisfaction. *Motiv Emot*. 2012;36(4):439–451.
12. Baumeister RF, Leary MR. The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychol Bull*. 1995;117(3):497–529.
13. Deci EL, Ryan RM. The “what” and “why” of goal pursuits: human needs and the self-determination of behavior. *Psychol Inq*. 2000;11(4):227–268.
14. Goldman N, Gleib DA, Seplaki C, Liu I-W, Weinstein M. Perceived stress and physiological dysregulation in older adults. *Stress*. 2005;8(2):95–105.
15. Franz CE, O’Brien RC, Hauger RL, et al. Cross-sectional and 35-year longitudinal assessment of salivary cortisol and cognitive functioning: the Vietnam Era Twin Study of Aging. *Psychoneuroendocrinology*. 2011;36(7):1040–1052.
16. Salsman JM, Lai JS, Hendrie HC, et al. Assessing psychological well-being: self-report instruments for the NIH toolbox. *Qual Life Res*. 2014;23(1):205–215.
17. Beaumont JL, Havlik R, Cook KF, et al. Norming plans for the NIH toolbox. *Neurology*. 2013;80(11 Suppl 3):S87–S92.
18. Casaletto KB, Umlauf A, Beaumont J, et al. Demographically corrected normative standards for the English version of the NIH toolbox cognition battery. *J Int Neuropsychol Soc*. 2015;21(5):378–391.
19. Reise S, Embretson S. *Item Response Theory for Psychologists*. Mahwah, New Jersey: Lawrence Erlbaum Associates. 2000.
20. Edwards Deming W, Stephan FF, Edwards Deming BW. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Source Ann Math Stat*. 1940; 11(4):427–444.
21. Battaglia MP, Izrael D, Hoaglin DC, Frankel MR. Practical considerations in raking survey data. *Surv Pract*. 2009;2(5):1–37.
22. Casaletto KB, Umlauf A, Marquine M, et al. Demographically corrected normative standards for the Spanish language version of the NIH toolbox cognition battery. *J Int Neuropsychol Soc*. 2016;22(3):364–374.
23. Rosseel Y. Lavaan: an R package for structural equation modeling. *J Stat Softw*. 2012;48(2):1–36.
24. Osborne JW, Costello AB. Best practices in exploratory factor analysis : four recommendations for getting the most from your analysis. *Pract Assess Res Eval*. 2005;10(7):1–9.
25. Tanaka, J. (1993). Multifaceted concepts of fit in structural equation models. In K. Bollen & S. Long (Eds.), *Testing Structural Equation Models* (10-39). Newberry Park CA: Sage.
26. Hoyle, R.H., & Panter, A.T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (158-176). Thousand Oaks, CA: Sage.
27. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull*. 1990;107(2):238–246.
28. Steiger JH. Structural model evaluation and modification: an interval estimation approach. *Multivariate Behav Res*. 1990;25(2):173–180.
29. MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychol Methods*. 1996;1(2):130–149.
30. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model A Multidiscip J*. 1999;6(1):1–55.
31. Brown T. (2015). *Confirmatory Factor Analysis for Applied Research*. Second Edition. New York City, NY: Guilford Press.
32. Taylor M, Heaton R. Sensitivity and specificity of WAIS-III/WMS-III demographically corrected factor scores in neuropsychological assessment. *J Int Neuropsychol Soc*. 2001;7(7):867–874.
33. Carlozzi NE, Goodnight S, Casaletto KB, et al. Validation of the NIH toolbox in individuals with neurologic disorders. *Arch Clin Neuropsychol*. 2017;32(5):555–573.
34. Siddiqi SH, Trapp NT, Hacker CD, et al. rTMS with individualized resting-state network mapping for neuropsychiatric sequelae of repetitive traumatic brain injury in a retired nfl player. *bioRxiv*. 2017. Available from: <https://www.biorxiv.org/content/early/2017/11/21/151696>. accessed February 19, 2018.
35. Health Measures. Available from: <http://www.healthmeasures.net/explore-measurement-systems/nih-toolbox/obtain-and-administer-measures>. Accessed February 19, 2018.

## Supplementary materials

**Table S1 Emotion Battery scales in factor solutions examined for best model fit**

<b>1-Factor model</b>	<b>2-Factor model</b>	<b>3-Factor model</b>	<b>4-Factor model</b>	<b>Revised 3-factor model</b>
Anger affect	<i>Negative affect</i>	<i>Negative affect</i>	<i>Negative affect</i>	<i>Negative affect</i>
Anger hostility	Anger affect	Anger affect	Anger affect	Anger affect
Anger physical aggression	Anger hostility	Anger hostility	Anger hostility	Anger hostility
Sadness	Anger physical aggression	Anger physical aggression	Anger physical aggression	Sadness
Fear affect	Sadness	Sadness	Sadness	Fear affect
Fear somatic arousal	Fear affect	Fear affect	Fear affect	Perceived stress
Life satisfaction	Fear somatic arousal	Fear somatic arousal	Fear somatic arousal	<i>Social satisfaction</i>
Meaning	Loneliness	Perceived stress	<i>Social relationships</i>	Friendship
Positive affect	Perceived hostility	<i>Social satisfaction</i>	Friendship	Loneliness
Friendship	Perceived stress	Friendship	Loneliness	Emotional support
Loneliness	Perceived rejection	Loneliness	Emotional support	Instrumental support
Emotional support	<i>Positive affect</i>	Emotional support	Instrumental support	Perceived rejection
Perceived hostility	Friendship	Instrumental support	Perceived rejection	<i>Psychological well-being</i>
Instrumental support	Emotional support	Perceived rejection	Perceived hostility	Life satisfaction
Perceived stress	Instrumental support	Perceived hostility	<i>Stress and self-efficacy</i>	Meaning
Perceived rejection	Meaning	<i>Psychological well-being</i>	Perceived stress	Positive affect
Self-efficacy	Life satisfaction	Life satisfaction	Self-efficacy	
	Self-efficacy	Meaning	<i>Psychological well-being</i>	
	Positive affect	Positive affect	Life satisfaction	
		Self-efficacy	Meaning	
			Positive affect	

**Note:** Factor names are shown in italics; scales that went into the factors are shown in non-italics.

**Table S2 Summary score formulas**

<b>Summary scores by language group</b>		<b>Mean</b>	<b>SD</b>
<b>English</b>			
Negative affect	$(((((a \times 0.774) + (b \times 0.644) + (c \times 0.842) + (d \times 0.827) + (e \times 0.835)) / 5) - (-0.02)) / 0.58) \times 10 + 50$	-0.02	0.58
Social satisfaction	$(((((i \times 0.709) + ((j \times -1) \times 0.804) + (k \text{ score} \times 0.760) + (l \times 0.586) + ((m \times -1) \times 0.703)) / 5) - (0.03)) / 0.53) \times 10 + 50$	0.03	0.53
Psychological well-being	$(((((f \times 0.752) + (g \times 0.729) + (h \times 0.879)) / 3) - (0.07)) / 0.67) \times 10 + 50$	0.07	0.67
<b>Spanish</b>			
Negative affect	$(((((a \times 0.749) + (b \times 0.602) + (c \times 0.860) + (d \times 0.879) + (e \times 0.750)) / 5) - (-0.01)) / 0.64) \times 10 + 50$	-0.01	0.64
Social satisfaction	$(((((i \times 0.620) + ((j \times -1) \times 0.753) + (k \times 0.761) + (l \times 0.670) + ((m \times -1) \times 0.757)) / 5) - (-0.08)) / 0.64) \times 10 + 50$	-0.08	0.64
Psychological well-being	$(((((f \times 0.783) + (g \times 0.794) + (h \times 0.864)) / 3) - (-0.05)) / 0.72) \times 10 + 50$	-0.05	0.72

**Notes:** Alphabet characters indicate scales designated in Tables S3 and S4. a, anger affect; b, anger hostility; c, sadness; d, fear affect; e, perceived stress; f, life satisfaction; g, meaning; h, positive affect; i, friendship; j, loneliness; k, emotional support; l, instrumental support; m, perceived rejection.

**Table S3** English-speaking raw scores conversion to standard scores

Scale	Formula
Anger affect	$[(X - (-0.10)) / 0.88] \times 10 + 50$
Anger hostility	$[(X - (-0.06)) / 0.93] \times 10 + 50$
Sadness	$[(X - (-0.04)) / 0.79] \times 10 + 50$
Fear affect	$[(X - (-0.16)) / 0.80] \times 10 + 50$
Perceived stress	$[(X - (-0.06)) / 0.99] \times 10 + 50$
Life satisfaction	$[(X - (-0.07)) / 0.92] \times 10 + 50$
Meaning	$[(X - (-0.11)) / 0.99] \times 10 + 50$
Positive affect	$[(X - (-0.05)) / 1.00] \times 10 + 50$
Friendship	$[(X - (-0.01)) / 0.96] \times 10 + 50$
Loneliness	$[(X - (-0.03)) / 0.94] \times 10 + 50$
Emotional support	$[(X - (-0.07)) / 0.95] \times 10 + 50$
Instrumental support	$[(X - (-0.05)) / 0.97] \times 10 + 50$
Perceived rejection	$[(X - (-0.04)) / 0.95] \times 10 + 50$
Self-efficacy	$[(X - (-0.05)) / 0.98] \times 10 + 50$
Anger-physical aggression	$[(X - (-0.04)) / 0.89] \times 10 + 50$
Fear somatic arousal	$[(X - (-0.04)) / 0.85] \times 10 + 50$
Perceived hostility	$[(X - (-0.04)) / 0.96] \times 10 + 50$

**Table S4** Spanish-speaking raw scores conversion to standard scores

Scale	Formula
Anger affect	$[(X - (-0.38)) / 1.16] \times 10 + 50$
Anger hostility	$[(X - (-0.02)) / 1.12] \times 10 + 50$
Sadness	$[(X - (-0.10)) / 1.17] \times 10 + 50$
Fear affect	$[(X - (-0.08)) / 1.16] \times 10 + 50$
Perceived stress	$[(X - (-0.11)) / 0.93] \times 10 + 50$
Life satisfaction	$[(X - (-0.08)) / 0.99] \times 10 + 50$
Meaning	$[(X - (-0.00)) / 0.98] \times 10 + 50$
Positive affect	$[(X - (-0.15)) / 0.92] \times 10 + 50$
Friendship	$[(X - (-0.31)) / 1.11] \times 10 + 50$
Loneliness	$[(X - (-0.01)) / 1.15] \times 10 + 50$
Emotional support	$[(X - (-0.21)) / 1.19] \times 10 + 50$
Instrumental support	$[(X - (-0.09)) / 1.09] \times 10 + 50$
Perceived rejection	$[(X - (-0.05)) / 1.15] \times 10 + 50$
<b>Supplemental scales</b>	
Self-efficacy	$[(X - (-0.25)) / 1.12] \times 10 + 50$
Anger-physical aggression	$[(X - (-0.07)) / 0.94] \times 10 + 50$
Fear somatic arousal	$[(X - (-0.20)) / 1.02] \times 10 + 50$
Perceived hostility	$[(X - (-0.37)) / 1.02] \times 10 + 50$

### Patient Related Outcome Measures

Dovepress

### Publish your work in this journal

Patient Related Outcome Measures is an international, peer-reviewed, open access journal focusing on treatment outcomes specifically relevant to patients. All aspects of patient care are addressed within the journal and practitioners from all disciplines are invited to submit their work as well as healthcare researchers and patient support groups.

The journal is included in PubMed. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/patient-related-outcome-measures-journal>