CrossMark

Measuring nonspecific factors in treatment: item banks that assess the healthcare experience and attitudes from the patient's perspective

Carol M. Greco[1,6] · Lan Yu[2] · Kelly L. Johnston[1] · Nathan E. Dodds[1] ·
Natalia E. Morone[2] · Ronald M. Glick[1,6] · Michael J. Schneider[3] ·
Mary Lou Klem[4] · Christine E. McFarland[1,6] · Suzanne Lawrence[1] ·
Jason Colditz[2] · Catherine C. Maihoefer[1] · Wayne B. Jonas[5] · Neal D. Ryan[1,6] ·
Paul A. Pilkonis[1]

Abstract

Purpose Nonspecific factors that accompany healthcare treatments, such as patients' attitudes and expectations, are important parts of the experience of care and can influence outcomes. However, no precise, concise, and generalizable instruments to measure these factors exist. We report on the development and calibration of new item banks, titled the Healing Encounters and Attitudes Lists (HEAL), that assess nonspecific factors across a broad range of treatments and conditions.

Methods The instrument development methodology of the Patient-Reported Outcomes Measurement Information System (PROMIS®) was used. Patient focus groups and clinician interviews informed our HEAL conceptual model. Literature searches of eight databases yielded over 500 instruments and resulted in an initial item pool of several thousand items. After qualitative item analysis, including cognitive interviewing, 296 items were included in field testing. The calibration sample included 1657 respondents, 1400 obtained through an Internet panel and 257 from conventional and integrative medicine clinics. Following exploratory and confirmatory factor analyses, the HEAL item banks were calibrated using item response theory.

Results The final HEAL item banks were Patient–Provider Connection (57 items), Healthcare Environment (25 items), Treatment Expectancy (27 items), Positive Outlook (27 items), and Spirituality (26 items). Short forms were also developed from each item bank. A six-item short form, Attitudes toward Complementary and Alternative Medicine (CAM), was also created.

Conclusions HEAL item banks provided substantial information across a broad range of each construct. HEAL item banks showed initial evidence of predictive and concurrent validity, suggesting that they are suitable for measuring nonspecific factors in treatment.

✉ Carol M. Greco
  grecocm@upmc.edu

1 Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

2 Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

3 Department of Physical Therapy, University of Pittsburgh School of Health and Rehabilitation Sciences, Pittsburgh, PA, USA

4 Health Sciences Library System, University of Pittsburgh, Pittsburgh, PA, USA

5 Samueli Institute, Alexandria, VA, USA

6 University of Pittsburgh Center for Integrative Medicine, 580 South Aiken Avenue, Suite 310, Pittsburgh, PA 15232, USA

Introduction

Two-key triple aim healthcare goals are enhancing the patient experience of care and improving outcomes [1]. Therefore, it is essential to assess patient perspectives of their Healthcare Environment and their personally relevant

outcomes. Although understanding the specific effects of treatments is vital and the focus of most biomedical research, nonspecific or contextual features of health care are also important as they contribute to the care experience and often influence outcomes [2]. The nonspecific, contextual features of healthcare treatments are broad ranging and multifaceted and may include interactive processes such as rituals and conditioning [3, 4], as well as patients' attitudes and their views of healthcare relationships and the treatment environment.

Trust in the healthcare provider, comfort with the clinical environment, and optimism regarding treatment are important contextual aspects of the patient experience of care. Such contextual factors may be independent of specific actions of a medication or procedure, yet there is growing recognition that they can influence treatment outcomes. A meta-analysis of 13 studies found significant effects of the patient–provider relationship on outcomes [5]. Trials that experimentally manipulate provider warmth show enhanced treatment effects even with inert treatments [6, 7]. Expectations regarding treatments are particularly influential in symptom improvement [6, 8–11]. For example, back pain patients with the highest treatment expectations exhibited an adjusted odds ratio of 5.3 for improvement [12]. Additionally, when given pills clearly marked 'placebo,' and told of potential mind–body benefits, 59 % of irritable bowel syndrome patients reported adequate improvement [13]. Likewise, patients' own characteristics shape their experience of care. Taking an active role can contribute to therapeutic outcomes [14], as can optimism [15, 16], and Spirituality [17]. A review of the biological mechanisms of nonspecific factors supports their influence on outcomes, often greater than the specific treatment [18]. Thus, enhancing nonspecific effects may be the most direct method for enhancing the impact of all therapies.

Efforts to understand nonspecific factors are increasing, particularly through experiments designed to elucidate placebo effects. However, there has been little effort to *measure* nonspecific, contextual factors, particularly from the patient's perspective. Instruments that measure these factors would have widespread value as they would allow researchers and clinicians to understand these dimensions as well as enhance them in the treatment setting.

This study reports on the development and initial validation of a set of patient-reported measures, the Healing Encounters and Attitudes Lists (HEAL), that assess several nonspecific factors in health care and healing. We used instrument development methodology of the Patient-Reported Outcomes Measurement Information System (PROMIS) [19, 20], which includes patient and clinician input, iterative steps of item revision, and modern psychometric methods (see Online Resource). HEAL

instruments are designed to be administered as computerized adaptive tests (CATs), in which a small number of items can provide precise information. However, brief static (six- to seven-item) forms were also created using this methodology. The University of Pittsburgh IRB approved this study (PRO13090474).

Methods

Overview

HEAL item banks were developed using the rigorous instrument development methodology of PROMIS [19, 21–23]. Several iterative steps were involved: formulation of a conceptual model, development of item banks, and psychometric analyses including both classical test theory (CTT) and item response theory (IRT).

### Formulation of the conceptual model

Multiple sources of information informed the model of contextual factors that contribute to health outcomes. We included the viewpoints of clinical experts and patients as well as existing scientific literature. Through interviews with over 20 healthcare providers, six patient focus groups that explored patients' ideas about contributors to positive and negative treatment outcomes, and literature reviews of placebo mechanisms and patient characteristics that moderate outcomes, we developed a preliminary conceptual model of HEAL (Fig. 1). The model included within-person characteristics and interpersonal relationship perceptions, and consisted of seven initial domains: positive/negative attitudes, Spirituality, Locus of Control, Treatment Expectancy, Health and Wellness Attitudes, perceptions of the patient–provider relationship, and patient perceptions of the Healthcare Environment.

Development of item banks

Item bank development included (1) comprehensive literature searches to find existing questionnaires, (2) categorization, review and refinement of existing items, and writing new items, and (3) cognitive interviews with patients to establish clarity (see Fig. 2).

Comprehensive literature searches

Searches of PubMed, EMBASE, CINAHL, AMED, ICL, MANTIS, HaPI, and PsycINFO databases were conducted using a search methodology developed by the Pittsburgh
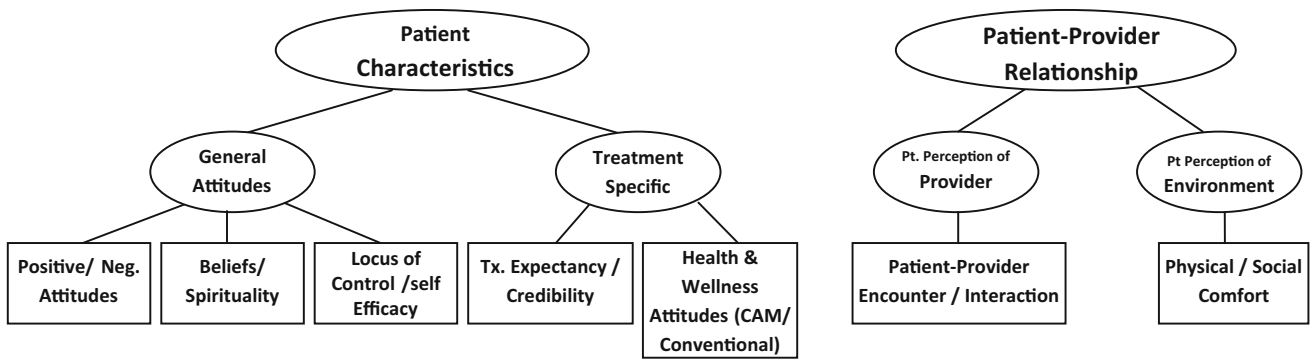
Fig. 1 Initial conceptual model of HEAL domains. Published originally as figure 1 in Greco et al. [25]

Fig. 2 Overview of HEAL item banks development
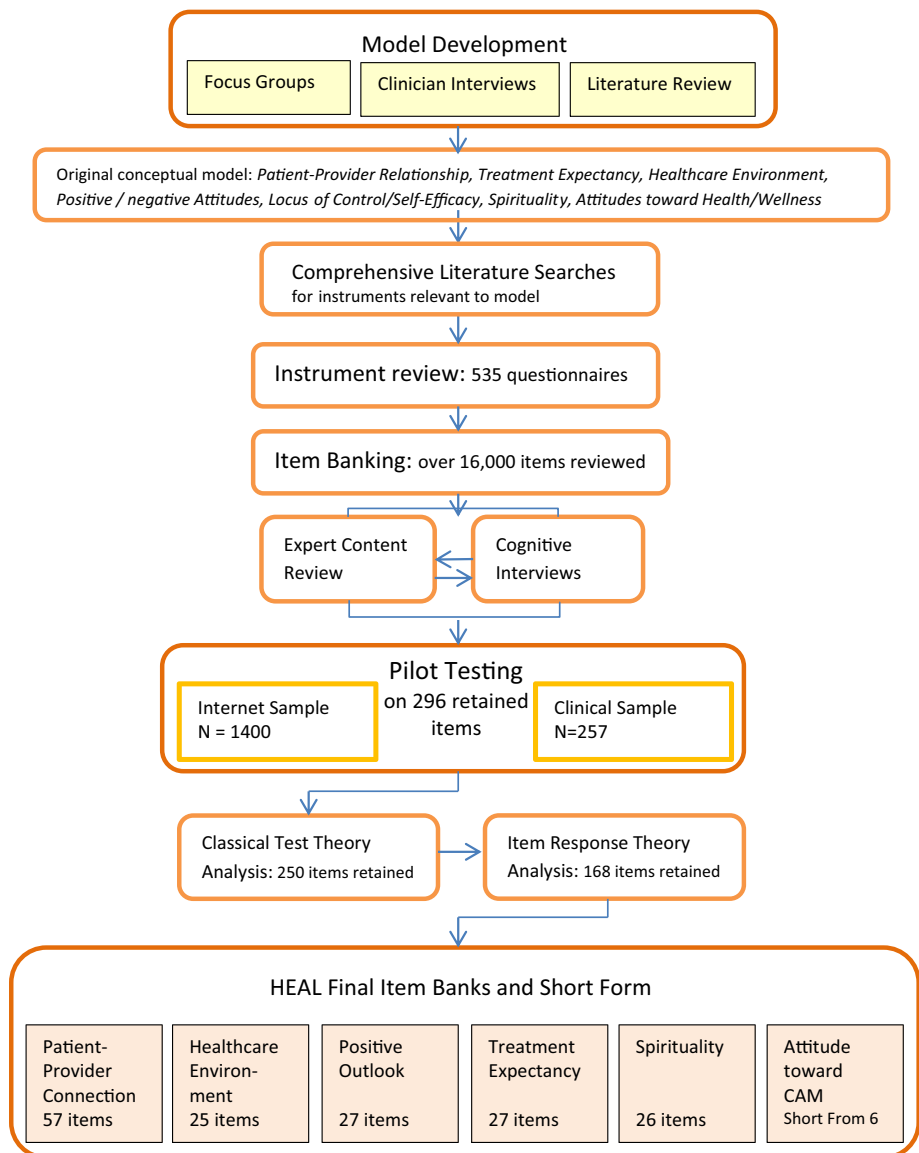
PROMIS research site [24]. The searches generated 14,864 abstracts, which were individually reviewed. Articles documenting development and psychometric validation of instruments were examined, yielding 535 unique questionnaires with over 16,000 items potentially related to the domains in the HEAL model.

Categorization and review of items

The initial pool of existing questionnaire items was coded into the conceptual categories of the HEAL model. The codings were based on independent ratings by teams of 2–4 healthcare clinician–researchers (R.G., C.G., N.M., M.J.S.) and instrument development experts (P.P., K.J., N.D., J.C., C.M.), with disagreements resolved through discussion. Using the qualitative item review procedures of PROMIS [19], the pool of items was reduced by removing redundant items and those that were too narrow in focus (e.g., disease or treatmentspecific). Further refinement included removing or rewriting poorly written items (see [25] for examples) and writing new items. The retained items were standardized for first-person subject, verb tense (primarily present tense), simple vocabulary, and five-point response scales reflecting frequency (*Never* to *Almost always*) or intensity (*Not at all* to *Very much*). From the initial pool of items, a total of 359 were retained for further review by patients in the cognitive interview stage.

Cognitive interviews

Forty-two patients participated in cognitive interviews, during which patients 'think aloud' while reviewing one item at a time with a trained interviewer. Each of the 359 items was reviewed by at least six patients representing a broad range of ages, diverse races and reading levels, and both genders. Patients provided feedback on the item clarity item, vocabulary, and appropriateness of the response scale. Based upon participant feedback, 63 items (17.5 %) were rewritten or removed. All revised items were subsequently reviewed in further cognitive interviews. The item pool contained 296 items following the cognitive interview phase.

Sampling

We field tested the items on two samples of patients who received conventional or integrative medicine treatments for a medical or mental health condition (details in Online Resource). One sample of 1400 persons was provided through the Internet survey company, YouGov.com. Patients who reported receiving treatments within the past year were eligible. The second sample included 257 patients at the University of Pittsburgh Medical Center (UPMC) who had recently started a new integrative medicine ($n = 127$) or conventional medicine ($n = 130$) treatment. This clinical sample repeated the computerized assessment 6 weeks later.

Measures

The 296 HEAL items retained following cognitive interviews represented several domains relevant to interpersonal and contextual aspects of treatment as well as potentially important patient attitudes.

To reduce burden for the Internet sample to 30 min or less, only HEAL items and demographics were included. Also, we administered subsets of HEAL items to respondents in a blocked design, ensuring that each item was paired with every other item a comparable number of times without requiring that participants respond to all items. In the clinical sample, in order to explore concurrent and discriminant validity, the PROMIS 29 health status profile and several well-known 'legacy' questionnaires were also completed (see Online Resource). At the 6-week follow-up, clinical participants also completed the single-item Clinical Global Impression (CGI) scale, which asks for comparison of current symptoms to earlier symptoms [26] to assess predictive validity of the HEAL measures.

Classical test theory analyses

The initial CTT analyses involved descriptive statistics and factor analysis. Although our conceptual model included seven domains deemed relevant to patients' experience of health care and healing, we made no assumptions regarding the most appropriate factor structure for the HEAL items. Our goal was to identify the most robust latent constructs with sufficient unidimensionality to proceed with IRT analyses. We also wanted to capture a variety of clinical indicators to ensure content validity. Thus, we intended to strike a balance between unidimensionality and adequacy of content.

The entire sample, both Internet and clinical, was randomly divided into two subsamples. The first subsample was used for exploratory factor analysis (EFA, $n = 799$), and the second was used for subsequent confirmatory factor analysis (CFA, $n = 858$) [27]. Factor loadings, scree plots, and eigenvalues were evaluated.

IRT analyses

Item response theory (IRT) refers to a set of psychometric methods for developing and scoring tests based upon the idea that one's response to each item reflects one's level on the underlying domain of interest. The latent trait (for example, Treatment Expectancy) is scaled along a dimension called theta (θ). IRT models include discrimination parameters (i.e., how well the item distinguishes among individuals higher vs. lower on the θ scale) and location/threshold parameters (i.e., the value of θ at which an individual has the highest probability of choosing the particular response to the item). Thus, IRT provides psychometric information about each item separately, as well as information for the overall test (see Online Resource).

Items remaining in the pools after CTT were calibrated with the two-parameter graded response model (GRM) using MULTILOG 7.03 [28]. Each item's fit to the IRT model was examined using the SAS macro-IRTFIT program [29]. Misfitting items (χ^2, $p < 0.01$) were considered for exclusion. Several additional outcomes from IRT analyses, such as differential item functioning (DIF) and indicators of local independence, were used to identify items for possible exclusion (see Online Resource).

Content-expert review

To ensure the relevance of the remaining items, content experts (C.G., P.P.) re-examined items from the clinical perspective to eliminate items with questionable psychometric properties or clinically overlapping content. Conversely, items with important clinical implications were considered for retention in the pool even if they did not meet some conventional psychometric guidelines.

Preliminary validity evidence

In order to evaluate concurrent and discriminant validity, we correlated θ scores on HEAL domains with corresponding legacy measures and the PROMIS health status measures. As an initial estimate of predictive validity, we correlated HEAL θ scores with participants' perceptions of improvement on the CGI. The clinical sample was the primary source for validity information, given the more extensive data we had on presenting complaints and treatment history.

Results

Factor analyses

In the EFA, we found a five-factor structure fit well to the data. Correlations among the five factors ranged from 0.17 to 0.49, indicating that they were related but sufficiently distinct to deserve separate status. The items in the original HEAL domains of Patient–Provider Connection and Healthcare Environment loaded on the first factor. The items in the original HEAL domains of Optimistic Attitudes and Locus of Control loaded on factor 2. Spirituality items, Health and Wellness Attitudes, and Treatment Expectancy items loaded primarily on factors 3, 4, and 5, respectively. Nine items with cross-loadings (0.40 loading on more than one factor) and 39 items with single-factor loadings <0.40 were dropped. Notably, 17 of the original 34 items in the Health and Wellness Attitudes item bank had low single-factor loadings (<0.40) and were among the items dropped. A second round of EFA was performed to

confirm the five-factor structure and to re-assess the magnitude of the factor loadings. No further items were eliminated after the second EFA.

Single-factor CFAs were performed for each of the five factors, based on items retained after EFAs. Iterations of CFA were performed until all retained items had loadings >0.50. A total of 250 items were retained at this stage for subsequent IRT analyses. To ensure construct validity, all items were reviewed again for content relevance. Based on this review, factor 1, which had 94 items, was separated into 1a, Patient–Provider Connection (PPC), and 1b, Healthcare Environment (HCE). CFA was rerun for PPC and HCE, and all factor loadings were larger than 0.50.

IRT calibrations

The item banks corresponding to the six factors from CTT were calibrated separately using the two-parameter GRM. Item-parameter estimates for the final items for each item bank are shown in the Online Resource in Tables 1–6.

Item information functions (IIF), differential item functioning, and local independence

Item information curves, which reflect the overall performance of individual items, were examined. Items with limited information (i.e., with peaks on the IIF < 1.0) were removed. Following IIF-based refinement, 190 items remained (see Table 1). No items showed differential functioning based upon gender, age, or education. Therefore, no items were removed due to DIF. Five locally dependent items were removed.

IRT model fit

We examined each item's fit to the IRT model. The SAS macro-IRTFIT identified 17 misfitting items, and these were removed (see Table 1).

The domain Health and Wellness Attitudes was originally broadly defined and included social and family factors, attitudes toward diet and exercise, fear of illness, and views of conventional and complementary medicine. The multidimensionality of this domain rendered 17 of the original 34 items inappropriate for fitting to IRT models. A further 11 items in this domain were dropped due to low item information (10 items) and IRT misfit (1 item), leaving too few items for CAT administration of this item bank. A set of six items concerning views of complementary medicine were the only survivors, and they were retained as a short, static measure entitled Attitudes toward CAM.

Table 1 Items retained following CTT and IRT analyses

Item bank	Domain	CFA		IIF		Local independence		IRTFIT	
		Factor	Retained	Removed	Retained	Removed	Retained	Removed	Retained
PPC	Patient–Provider Connection	1a	61	2	59	0	59	2	57
HCE	Healthcare Environment	1b	33	6	27	0	27	2	25
TE	Treatment Expectancy	5	42	1	41	3	38	11	27
PO	Positive Outlook	2	64	37	27	0	27	0	27
SP	Spirituality	3	33	4	29	2	27	1	26
HWA	Health and Wellness Attitudes	4	17	10	7	0	7	1	6 (Attitudes toward CAM)
			250	60	190	5	185	17	168

Preliminary validity evidence

In the clinical sample, we examined concurrent validity between the six HEAL θ scores and legacy instruments measuring similar constructs. HEAL Patient–Provider Connection and Healthcare Environment correlated 0.38 and 0.39 ($p < 0.01$) with a measure of outpatient clinical care (ACES [30]), indicating similarity but not complete overlap. HEAL Treatment Expectancy was associated with Credibility 0.71 ($p < 0.01$) and Expectancy 0.58 ($p < 0.01$) factors of the Credibility Expectancy Questionnaire [31]. HEAL Positive Outlook was inversely associated with PROMIS 29 depression (-0.71, $p < 0.01$) and anxiety (-0.54, $p < 0.01$). Likewise, Spirituality, Positive Outlook, and Attitudes toward CAM were significantly correlated with corresponding legacy measures (see Online Resource), with correlations ranging from 0.81 to 0.30 (all p's < 0.01). Treatment-related item banks PPC and HCE were unrelated to PROMIS health status measures, providing preliminary support for discriminant validity.

To explore predictive validity, baseline HEAL scores were compared with follow-up CGI ratings in the clinical sample. Correlations ranged from 0.36 ($p < 0.01$) for HEAL Treatment Expectancy to 0.13, ($p < 0.05$) for HEAL Spirituality, indicating that HEAL scores account for some variability in patients' perceived improvement across a broad range of treatments.

Selection of items for short forms

To administer the HEAL where computerized adaptive testing (CAT) is not available, static short forms can be a useful alternative. To develop short forms, we evaluated the items in the HEAL banks based on their psychometric properties: their discrimination parameters, the percentage

of time the item would have been selected in a simulated CAT [32] based on the observed data from the calibration samples, the expected information under the standard normal distribution (mean 0, SD 1), and the expected information under an extended distribution (mean 0, SD 1.5) [33]. For five of the HEAL banks, we selected items for short forms based upon the convergence of the psychometric criteria together with review of the clinical importance and content balance of the items. Clinical experts (P.P., C.G.) performed these content reviews and choose items for short forms that had both excellent psychometric properties and represented the clinical breadth of the domains. The other HEAL item bank, originally conceptualized as Attitudes toward Health and Wellness, included only six items following CTT and IRT analyses which is an inadequate number for CAT. Therefore, this bank is available only as a short form: Attitudes toward Complementary/Alternative Medicine (CAM). The short forms are provided in Table 2.

The short forms' internal consistency alpha coefficients were 0.92 for Healthcare Environment and Positive Outlook, 0.96 for Patient–Provider Connection and Treatment Expectancy, and 0.97 for Spirituality. The correlations between the theta scores derived from the short forms and their corresponding full item banks were high, ranging from 0.93 to 0.97, indicating that the short form scores are highly consistent with those of the full banks.

Discussion

Nonspecific or contextual aspects of healthcare treatments are complex and extensive and contribute to patients' experience of care. Nonspecific factors may moderate and mediate treatment outcomes, and thus, despite the challenges, are important to assess in medical research. Barriers

Table 2 Calibrated HEAL short form items

Patient–Provider Connection (PPC)	Slope	Location thresholds			
Think of the HCP (Healthcare Provider) who provides your current/ongoing treatment…	(Discrimination)	Not at all versus a little bit	A little bit versus somewhat	Somewhat versus quite a bit	Quite a bit versus very much
I am satisfied with my healthcare provider	4.29	−1.41	−0.88	−0.09	0.71
I trust my healthcare provider	3.93	−1.66	−0.95	−0.19	0.74
My healthcare provider pays attention to my individual needs	3.80	−1.36	−0.85	0.08	0.89
My healthcare provider gives me enough information[a]	3.57	−1.82	−0.98	0.00	0.95
My healthcare provider respects me	3.55	−1.81	−1.10	−0.23	0.71
I feel my healthcare provider understands me	3.52	−1.22	−0.68	0.19	1.00
My healthcare provider gives me support and encouragement	3.51	−1.41	−0.80	0.07	0.95

Healthcare Environment (HCE)	Slope	Location threshold			
Think of the place where you receive your current/ongoing treatment…	(Discrimination)	Not at all versus a little bit	A little bit versus somewhat	Somewhat versus quite a bit	Quite a bit versus very much
The staff was respectful	4.19	−2.75	−1.92	−1.17	−0.29
The staff was friendly	3.91	−2.60	−1.75	−1.13	−0.21
The staff was helpful	3.73	−2.56	−1.91	−1.05	−0.14
My care was well organized	3.09	−2.54	−1.91	−0.98	−0.04
The healthcare provider's office respected my privacy	3.08	−2.51	−2.00	−1.25	−0.28
The waiting area was comfortable	2.35	−2.70	−1.98	−0.83	0.15

Treatment Expectancy (TE)	Slope	Location threshold			
	(Discrimination)	Not at all versus a little bit	A little bit versus somewhat	Somewhat versus quite a bit	Quite a bit versus very much
I am confident in this treatment	4.64	−1.56	−0.90	−0.03	0.86
This treatment will be successful	4.34	−1.71	−0.87	0.06	0.99
I feel good about this treatment	4.25	−1.60	−0.98	−0.15	0.74
I expect good outcomes from this treatment	4.23	−1.68	−0.86	−0.02	0.85
This treatment is right for me	4.02	−1.68	−1.06	−0.14	0.80
I value this treatment	3.47	−1.67	−1.15	−0.25	0.67

Positive Outlook (PO)	Slope	Location threshold			
	(Discrimination)	Not at all versus a little bit	A little bit versus somewhat	Somewhat versus quite a bit	Quite a bit versus very much
I feel positive about my life	4.25	−1.95	−1.18	−0.30	0.68
I am hopeful about my future	3.03	−2.06	−1.15	−0.30	0.59
My future looks good	2.95	−1.92	−1.19	−0.11	0.85
I am satisfied with my life	2.82	−1.81	−1.12	−0.13	1.01
I feel confident about myself	2.56	−2.11	−1.16	−0.21	0.91
I feel I can cope with my problems	2.09	−2.72	−1.63	−0.46	0.98

Table 2 continued

Spirituality (Sp)	Slope	Location threshold			
	(Discrimination)	Not at all versus a little bit	A little bit versus somewhat	Somewhat versus quite a bit	Quite a bit versus very much
Spiritual beliefs give meaning to my life	4.29	−1.04	−0.42	0.14	0.73
Spiritual beliefs give me hope	4.29	−1.13	−0.49	0.09	0.64
I find comfort in my faith	4.15	−0.98	−0.42	0.10	0.70
My Spirituality gives me inner strength	3.89	−1.22	−0.46	0.10	0.76
Prayer is a meaningful part of my life	3.87	−0.78	−0.26	0.23	0.71
I feel supported by a higher power	3.73	−1.03	−0.47	0.06	0.63

Attitudes toward CAM	Slope	Location threshold			
CAM (Complementary and Alternative Medicine) is a non-conventional, holistic, or natural approach to health care. Common CAM treatments may include acupuncture, massage therapy, meditation, or herbal remedies	(Discrimination)	Not at all versus a little bit	A little bit versus somewhat	Somewhat versus quite a bit	Quite a bit versus very much
CAM is effective	3.18	−2.20	−0.83	0.39	1.27
I prefer CAM over conventional medicine	2.80	−0.49	0.15	0.91	1.55
It is important to be open to CAM	2.65	−2.06	−1.00	−0.12	0.65
CAM can be used to treat serious illness	2.43	−1.51	−0.43	0.60	1.35
CAM can prevent health problems	2.27	−1.99	−0.78	0.46	1.39
I prefer natural remedies	2.15	−1.38	−0.33	0.57	1.31

[a] Response categories are: Never, Rarely, Sometimes, Often, Almost always

to such assessments include patient and clinician burden and a lack of precise measures that are relevant across a broad range of treatments and health conditions. Many existing measures of patient engagement and other relevant factors are lengthy, specific to a particular setting, or are not informative across the full range of the construct. We developed the HEAL item banks to overcome these barriers and provide researchers and clinicians with precise tools for assessing contextual factors that influence treatment outcomes and patients' care experiences.

The HEAL banks of Patient–Provider Connection (57 items), Healthcare Environment (25 items), Treatment Expectancy (27 items), Positive Outlook (27 items), and Spirituality (26 items) demonstrated adequate unidimensionality, high item and test information, and initial construct validity. The items of our original model domains of Locus of Control and Self-Efficacy were not found to be separate factors distinct from Positive Outlook, and thus, these items are contained within the Positive Outlook item bank. The domain Attitudes toward Health and Wellness originally included many items describing health attitudes ranging from views of exercise to opinions regarding the role of advanced technology in health care. Due to this diversity, upon psychometric testing very few of the items

in this domain formed a unidimensional set of items. This set of items concerned Attitudes toward CAM and were redefined as a short six-item form. We were able to derive short forms for the other five HEAL banks that provide brief, efficient measures of nonspecific moderating and mediating factors that may contribute to healing outcomes. Short forms may be especially useful in clinic to assess the contribution of nonspecific factors to patients' experiences and outcomes.

HEAL CATs and associated short forms fill an important gap in the measurement of nonspecific factors that may affect treatment outcomes. The HEAL are unique in that they provide precise information on potentially influential interpersonal and attitudinal factors that cut broadly across medical conditions and treatments. Because they are administered as CATs or short forms, using HEAL does not unduly burden patients. Investigators and clinicians can choose to use whichever HEAL instruments are most relevant and appropriate for their purposes. In research, the use of HEAL CATs may be a step toward dismantling the 'black box' of the placebo effect. In clinical applications, HEAL may inform clinicians regarding factors that can enhance health outcomes and improve the experience of care [34].

Future directions

One goal is to provide further validation of the HEAL item banks among patients who are initiating a treatment that involves ongoing meetings with a healthcare provider. Our initial validation work in the present sample provides useful clues to predictive validity of the HEAL item banks, yet is limited by the fact that treatments were of differing lengths and intensity, and clinical details were limited. We are currently validating the HEAL CATs and short forms in patients with chronic back or neck pain receiving new treatments. A related goal is to determine, through both qualitative and quantitative methods, the clinical utility of the HEAL for improving patient–provider engagement and clinical outcomes in an expanded group of patients. Future work to disseminate and implement HEAL as well as other PROMIS measures will involve their systematic inclusion in large behavioral and pharmacologic RCTs.

Compliance with ethical standards

Conflict of interest   All authors declare that they have no conflicts of interests regarding the study.

Ethical approval   All procedures performed were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the Institutional Review Board of the University of Pittsburgh.

Informed consent   Informed consent was obtained from all focus group participants, cognitive interview participants, and clinical calibration participants.

References

1. Berwick, D. M., Nolan, T. W., & Whittington, J. (2008). The triple aim: Care, health, and cost. *Health Affairs, 27*(3), 759–769.
2. *National Center for Complementary and Alternative Medicine Third Strategic Plan 2011–2015*. Exploring the science of complementary and alternative medicine 2011. http://nccam.nih.gov/sites/nccam.nih.gov/files/about/plans/2011/NCCAM_SP_508.pdf.
3. Enck, P., Benedetti, F., & Schedlowski, M. (2008). New insights into the placebo and nocebo responses. *Neuron, 59*(2), 195–206.
4. Benedetti, F., Amanzio, M., Benedetti, F., & Amanzio, M. (2011). The placebo response: how words and rituals change the patient's brain. *Patient Education and Counseling, 84*(3), 413–419.
5. Kelley, J. M., Kraft-Todd, G., Schapira, L., Kossowsky, J., & Riess, H. (2014). The influence of the patient–clinician relationship on healthcare outcomes: A systematic review and meta-analysis of randomized controlled trials. *PLoS One, 9*(4), e94207.
6. Kaptchuk, T. J., Kelley, J. M., Conboy, L. A., Davis, R. B., Kerr, C. E., Jacobson, E. E., et al. (2008). Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. *BMJ, 336*(7651), 999–1003.
7. Rakel, D., Barrett, B., Zhang, Z., Hoeft, T., Chewning, B., Marchand, L., et al. (2011). Perception of empathy in the therapeutic encounter: Effects on the common cold. *Patient Education and Counseling, 85*(3), 390–397.
8. Diederich, N. J., & Goetz, C. G. (2008). The placebo treatments in neurosciences: New insights from clinical and neuroimaging studies. *Neurology, 71*(9), 677–684.
9. Grenfell, R. F., Briggs, A. H., & Holland, W. C. (1961). A double-blind study of the treatment of hypertension. *JAMA, 176*(2), 124–130.
10. Wasan, A. D., Kong, J., Pham, L. D., Kaptchuk, T. J., Edwards, R., & Gollub, R. L. (2010). The impact of placebo, psychopathology, and expectations on the response to acupuncture needling in patients with chronic low back pain. *Journal of Pain, 11*(6), 555–563.
11. Vase, L., Baram, S., Takakura, N., Yajima, H., Takayama, M., Kaptchuk, T. J., et al. (2013). Specifying the nonspecific components of acupuncture analgesia. *Pain, 154*(9), 1659–1667.
12. Kalauokalani, D., Cherkin, D. C., Sherman, K. J., Koepsell, T. D., & Deyo, R. A. (2001). Lessons from a trial of acupuncture and massage for low back pain. *Spine, 26*, 1418–1424.
13. Kaptchuk, T. J., Friedlander, E., Kelley, J. M., Sanchex, M. N., Kokkotou, E., Singer, J. P., et al. (2010). Placebos without deception: A randomized controlled trial in irritable bowel syndrome. *PLoS One, 5*(10), 1–7.
14. Barrett, B., Muller, D., Rakel, D., Rabago, D., Marchand, L., & Scheder, J. C. (2006). Placebo, meaning, and health. *Perspectives in Biology and Medicine, 49*(2), 178–198.
15. Walach, H., & Jonas, W. B. (2004). Placebo research: the evidence base for harnessing self-healing capacities. *Journal of Alternative and Complementary Medicine, 10*(Suppl 1), S103–S112.
16. Price, D. D., Finniss, D. G., & Benedetti, F. (2008). A comprehensive review of the placebo effect: recent advances and current thought. *Annual Review of Psychology, 59*, 565–590.
17. Hyland, M. E., Geraghty, A. W. A., Joy, O. E. T., & Turner, S. I. (2006). Spirituality predicts outcome independently of expectancy following flower essence self-treatment. *Journal of Psychosomatic Research, 60*, 53–58.
18. Meissner, K., Kohls, N., & Colloca, L. (2011). Introduction to placebo effects in medicine: mechanisms and clinical implications. *Philosophical Transactions of the Royal Society of London-Series B: Biological Sciences, 366*, 1783–1789.
19. Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., et al. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Medical Care, 45*(5 Suppl 1), S3–S11.
20. 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*(Suppl 1), 133–141.
21. Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., et al. (2010). The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks:

2005–2008. *Journal of Clinical Epidemiology, 63*(11), 1179–1194.

22. DeWalt, D. A., Rothrock, N., Yount, S., Stone, A. A., & PROMIS Cooperative Group. (2007). Evaluation of item candidates: the PROMIS qualitative item review. *Medical Care, 45*(5 Suppl 1), S12–S21.

23. Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., et al. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care, 45*(5 Suppl 1), S22–S31.

24. Klem, M. L., Saghafi, E., Abromitis, R., Stover, A., Dew, M. A., & Pilkonis, P. (2009). Building PROMIS item banks: Librarians as co-investigators. *Quality of Life Research, 18*, 881–888.

25. Greco, C. M., Glick R. M., Morone N. E., & Schneider M. J. (2013) Addressing the "It is just placebo" pitfall in CAM: Methodology of a project to develop patient-reported measures of nonspecific factors in healing. *Evidence-Based Complementary and Alternative Medicine, 2013,* 613797. doi:10.1155/2013/613797.

26. Guy, W. (1976). *Clinical global impressions: ECDEU assessment manual for psychopharmacology* (pp. 218–222). Rockville, MD: National Institute of Mental Health.

27. Muthen, L. K., & Multhen, B. O. (2007). *Mplus user's guide* (4th ed.). Mullen & Mullen: Los Angeles, CA.

28. Thissen, D., Chen, W.-H., & Bock, R. D. (2003). *Multilog (version 7)*. Lincolnwood, IL: Scientific Software International.

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

30. Safran, D. G., Karp, M., Coltin, K., Chang, H., Li, A., Ogren, J., & Rogers, W. H. (2006). Measuring patients' experiences with individual primary care physicians. Results of a statewide demonstration project. *Journal of General Internal Medicine, 21*(1), 13–21.

31. Devilly, G. J., & Borkovec, T. D. (2000). Psychometric properties of the credibility/expectancy questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry, 31*(2), 73–86.

32. Choi, S. W. (2009). Firestar: Computerized adaptive testing (CAT) simulation for polytomous IRT models. *Applied Psychological Measurement, 33*, 644.

33. Choi, S. W., Reise, S. P., Pilkonis, P. A., Hays, R. D., & Cella, D. (2010). Efficiency of static and computer adaptive short forms compared to full length measures of depressive symptoms. *Quality of Life Research, 19*, 125–135.

34. *Institute for Healthcare Improvement*. www.ihi.org.