Title
Development of the PROMIS health expectancies of smoking item banks.

Journal
Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco, 16 Suppl 3(Suppl 3)

ISSN
1462-2203

Authors
Edelen, Maria Orlando
Tucker, Joan S
Shadel, William G
et al.

Publication Date
2014-09-01

DOI
10.1093/ntr/ntu053

Peer reviewed

Nicotine & Tobacco Research, Volume 16, Number 3 (August 2014) S223–S231

ORIGINAL INVESTIGATION

Development of the PROMIS® Health Expectancies of Smoking Item Banks

Maria Orlando Edelen PhD[1], **Joan S. Tucker** PhD[2], **William G. Shadel** PhD[3], **Brian D. Stucky** PhD[2], **Jennifer Cerully** PhD[3], **Zhen Li** MS[4], **Mark Hansen** MPH[4], **Li Cai** PhD[4]

[1]*RAND Health, RAND Corporation, Boston, MA;* [2]*RAND Health, RAND Corporation, Santa Monica, CA;* [3]*RAND Health, RAND Corporation, Pittsburgh, PA;* [4]*CSE/CRESST, Graduate School of Education and Information Studies, University of California, Los Angeles, Los Angeles, CA*

Corresponding Author: Maria Orlando Edelen, PhD, RAND Health, RAND Corporation, 20 Park Plaza, Suite 920, Boston, MA 02116, USA. Telephone: 617-338-2059, ext. 8634; Fax: 617-357-7470; E-mail: orlando@rand.org

Received February 24, 2014; accepted March 9, 2014

ABSTRACT

Introduction: Smokers' health-related outcome expectancies are associated with a number of important constructs in smoking research, yet there are no measures currently available that focus exclusively on this domain. This paper describes the development and evaluation of item banks for assessing the health expectancies of smoking.

Methods: Using data from a sample of daily (*N* = 4,201) and nondaily (*N* = 1,183) smokers, we conducted a series of item factor analyses, item response theory analyses, and differential item functioning analyses (according to gender, age, and race/ethnicity) to arrive at a unidimensional set of health expectancies items for daily and nondaily smokers. We also evaluated the performance of short forms (SFs) and computer adaptive tests (CATs) to efficiently assess health expectancies.

Results: A total of 24 items were included in the Health Expectancies item banks; 13 items are common across daily and nondaily smokers, 6 are unique to daily, and 5 are unique to nondaily. For both daily and nondaily smokers, the Health Expectancies item banks are unidimensional, reliable (reliability = 0.95 and 0.96, respectively), and perform similarly across gender, age, and race/ethnicity groups. A SF common to daily and nondaily smokers consists of 6 items (reliability = 0.87). Results from simulated CATs showed that health expectancies can be assessed with good precision with an average of 5–6 items adaptively selected from the item banks.

Conclusions: Health expectancies of smoking can be assessed on the basis of these item banks via SFs, CATs, or through a tailored set of items selected for a specific research purpose.

INTRODUCTION

The PROMIS® Smoking Initiative is developing, evaluating, and making available a set of psychometrically sound item banks that can form the basis for standardized assessment of cigarette smoking behavior and biopsychosocial constructs associated with smoking. Our state-of-the-art approach utilizes item banking via item response theory (IRT), which results in an extremely versatile and sustainable assessment system that maximizes measurement precision while minimizing respondent burden (Edelen & Reeve, 2007; Embretson & Reise, 2000). This paper describes the development and evaluation of the Health Expectancies of Smoking item banks for daily and nondaily smokers. Our development was guided by a conceptual framework of smoking behavior and related constructs, which included a domain capturing the health-related aspects of smoking (Edelen, Tucker, Shadel, Stucky, & Cai, 2012). Using this conceptual framework, we conducted a systematic literature review and a series of focus group discussions with smokers which led to the identification of over 1,500 smoking items. A subset of these items was selected for inclusion in an initial item pool and administered to over 3,000 daily and nondaily smokers in the United States. Data from the field test were used to conduct extensive quantitative analyses to identify salient and distinct smoking-related domains (Edelen et al., 2012).

One of the domains that emerged as a result of this process was comprised of items related to outcome expectancies regarding the physical health effects of smoking. Some of the items' content covered health-related outcome expectancies of smoking, such as current health-related consequences of smoking (e.g., "Smoking causes me to get tired easily") and concerns about the effects that smoking will have on future health and well-being (e.g., "I worry that smoking will lower my quality of life"). Other items' content covered health-related outcome expectancies of quitting, such as anticipated health improvements that could be realized upon quitting smoking (e.g., "If I quit smoking I will be healthier"). We use the label

doi:10.1093/ntr/ntu053

Health Expectancies of Smoking (heretofore referred to as Health Expectancies) to characterize this set of items.

Outcome expectancies play a central role in social learning theory models of substance use (Bandura, 1986), and health-related outcome expectancies are particularly salient for smokers, as the association between smoking and health consequences is well established (U.S. Department of Health and Human Services, 2004). In addition, health-related outcome expectancies (or perceived health risks and benefits) are also fundamental constructs of many health behavior theories (Brewer et al., 2007; Weinstein, 1993), and smokers' health-related outcome expectancies of smoking and quitting are associated with a number of smoking-related constructs, such as intentions to smoke, decisions to start smoking, decisions to quit, and successful periods of abstinence (McKee, O'Malley, Salovey, Krishnan-Sarin, & Mazure, 2005; Romer & Jamieson, 2001; Weinstein, 2001). Many smoking-related intervention strategies capitalize on the relationship between health-related outcome expectancies and smoking behavior (Bize et al., 2012; Schlam & Baker, 2013). A common approach in public health interventions targeting smoking prevention or cessation is use of fear appeals to increase perceptions of health threat and downstream to change smoking behavior (Witte & Allen, 2000). For example, policies requiring graphic labels on cigarette packs and the implementation of media campaigns featuring vivid examples of the negative health outcomes of smoking utilize fear appeals (Strasser, Tang, Romer, Jepson, & Cappella, 2012). Other interventions are educational in nature and give smokers accurate information about possible negative health consequences of smoking (e.g., risk of developing lung cancer or cardiovascular disease) in the hopes of changing smokers' negative health-related outcome expectancies and leading to better decision making about smoking (Brown, 2003). Further, doctors' advice about the personal health effects associated with smoking influences their patients' quitting (Stead et al., 2013).

Health-related outcome expectancies of smoking and quitting have been assessed as part of several validated scales. Two instruments designed specifically to assess outcome expectancies for smoking are the Smoking Effects Questionnaire (SEQ; Rohsenow et al., 2003) and the Smoking Consequences Questionnaire (SCQ; Brandon & Baker, 1991; Copeland, Brandon, & Quinn, 1995). The SEQ has two second-order factors assessing health-related outcome expectancies including "negative physical effects" and "future health concerns." The physical effects factor includes items such as "Smoking makes me feel weaker physically," and the future health concerns factor is measured with statements like "Smoking makes me worry about getting or having cancer." Similarly, within its factor assessing negative consequences of smoking, the four-factor SCQ includes 10 health-related outcome expectancy items (this factor also includes items reflecting addiction sustainment and negative social impression). The 10 health-related items ask smokers to rate their agreement with statements related to health risk (five items, e.g., "Smoking is taking years off my life") and respiratory irritation (five items, e.g., "Cigarettes make my lungs hurt"). Additionally, the 39-item Perceived Risks and Benefits Questionnaire (PRBQ; McKee et al., 2005) assesses outcome expectancies (risks and benefits) associated with quitting smoking and includes items reflecting health benefits (five items, e.g., "If I quit smoking I will live longer") and general well-being (four items, e.g., "If I quit smoking I will feel more energetic").

In short, health-related outcome expectancies of smoking are theoretically important, are associated with multiple important smoking-related outcomes, and are a key part of many approaches to prevention and treatment options for smokers. However, these expectancies are measured across several instruments, and each one emphasizes slightly different aspects of the domain content (e.g., smoking vs. quitting). There are currently no measures that exclusively assess health-related outcome expectancies of smoking and quitting or that cover the breadth of content represented by the collection of items we identified. This underscores the importance of developing a validated set of items to measure this construct.

In this paper, we first describe the series of analyses we conducted to arrive at a unidimensional set of items assessing the health-related expectancies of smoking among daily and nondaily smokers. We then describe how we developed and evaluated the performance of short forms and computer adaptive tests (CATs) to efficiently, yet reliably assess this domain. Our analysis plan follows closely the procedures described by Reeve et al. (2007) to psychometrically evaluate and calibrate health-related quality of life item banks for PROMIS. More details of the analytic process used to develop the daily and nondaily smoker Health Expectancies item banks can be found in Hansen et al. in this supplement.

METHODS

Sample and Procedure

A national sample of smokers ($N_{(total)}$ = 5,384; $N_{(daily)}$ = 4,201; $N_{(nondaily)}$ = 1,183) was recruited by Harris Interactive through their online panel membership, and all assessments were completed via the internet. This total (5,384) represents 94% of eligible respondents (total eligible = 5,735); the remaining 6% (n = 351) started the survey but suspended early and were not included in the final data set. All procedures were IRB approved. Individuals were eligible if they were 18 years or older, had been smoking for at least a year, had smoked in the past 30 days, and did not have plans to quit in the next 30 days. Based on their response to number of days smoked in past 30 days, those participants indicating smoking 28–30 of the past 30 days were classified as daily smokers; respondents smoking less than 28 of the past 30 days were classified as nondaily smokers. Sample recruitment was targeted to reflect the demographic composition of U.S. adult smokers in terms of gender, race/ethnicity, and age. The survey was fielded between July and September 2011 via a randomized block design (Reeve et al., 2007). The block design was constructed to minimize respondent burden while maximizing the inter-item covariance coverage. To cross-validate the dimensionality of the Health Expectancies item bank, the daily smoker sample was randomly split into exploratory ($N_{(exploratory)}$ = 3,021) and confirmatory ($N_{(confirmatory)}$ = 1,180) subsamples.

Mean age was 46.4 years for daily (D) smokers and 44.1 years for nondaily (ND) smokers. Females comprised about half the sample (D: 54.8%, ND: 47.0%). Most participants were employed full-time (D: 52.9%, ND: 60.6%) or part-time (D: 12.2%, ND: 14.4%). The racial/ethnic composition was primarily non-Hispanic White (D: 72.2%, ND: 55.2%), African American (D: 12.1%, ND: 15.5%), and Hispanic (D: 11.3%, ND: 24.4%). Most participants had attended at least some college (D: 80.5%, ND: 84%), and many had earned a bachelors or graduate degree (D: 29.8%, ND: 42.1%). More than half were currently married or cohabitating (D: 57.7%, ND: 55.1%), with fewer being divorced/separated/widowed (D: 21.8%, ND: 18.7%) or never married (D: 20.5%, ND: 26.1%). Although most differences are not large, chi-square tests (and *t*-test for age) indicated that daily and nondaily smokers significantly differed on each of these characteristics ($p < .001$). Most notably, relative to daily smokers, nondaily smokers were less likely to be non-Hispanic White, and more likely to be employed and further educated. Table 1 compares these groups on smoking patterns. As expected, daily smokers had a longer smoking history, smoked more heavily, and reported fewer quit attempts compared with nondaily smokers ($p < .0001$).

Measures

Smoking Items

A total of 277 unique smoking items were administered. These items were developed according to PROMIS procedures from extant items in the literature as well as direct feedback from smokers. This process, described in more detail in Edelen et al. (2012), employed a rigorous qualitative approach that included systematic literature review, categorization of items into domain "bins," removal of items with redundant or irrelevant content or unwieldy wording (e.g., binning and winnowing; DeWalt, Rothrock, Yount, & Stone, 2007), item standardization so that all the items were presented in a similar format with standard response options, solicitation of feedback from over 100 smokers via focus groups (4 groups with 38 total participants) and cognitive interviews (with 66 smokers), and final item revisions. Thirteen of the 277 smoking items which assessed smoking behavior and quitting history were completed by all field test respondents. The remaining 264 items were candidate items that were being considered for inclusion in one of the smoking item banks. These items were distributed across 26 overlapping forms containing an average of 147 items (range = 134–158); each respondent was randomly assigned one of the 26 forms.

Other Measures

All respondents supplied basic demographic information and completed one of eight PROMIS health-related quality of life short-form measures (alcohol consumption, anger, anxiety, depression, fatigue, physical functioning, sleep disturbance, and global health; Cella et al., 2007). These PROMIS measures were collected to provide preliminary validity evidence and results are reported elsewhere in this supplement.

Item Factor Analyses

Previous analyses of the daily smoker exploratory subsample identified a set of 26 items to be considered for inclusion in the Health Expectancies item bank for daily smokers (Edelen et al., 2012). These items included content based on existing items in the SCQ, the SEQ, and the PRBQ, as well as content that emerged from focus group discussions (e.g., "It takes me longer to recover from a cold because I smoke"). The same 26 items were also considered for nondaily smokers.

Table 1. Smoking Characteristics of Daily and Nondaily Smokers

Smoking variable	Daily smokers (N = 4,201)	Nondaily smokers (N = 1,183)
Years smoked, %		
1–10 years	11.7	29.2
>10 years	88.3	70.8
Number of days smoked in past 30, %		
1 or 2 days	0.0	15.8
3–5 days	0.0	9.6
6–9 days	0.0	9.6
10–19 days	0.0	23.2
20–27 days	0.0	41.9
28–30 days	100.0	0.0
Average number of cigarettes per day in past 30 days, %		
<1 per day	0.2	13.0
1–5	8.0	48.3
6–10	22.0	22.3
11–20	47.3	13.5
20+	22.6	3.9
Number of times quit for at least 24 hr, %		
Never	18.0	14.7
1 time	12.3	6.2
2–3 times	30.7	19.1
4–5 times	19.7	12.7
6–9 times	7.4	7.8
10 or more times	12.0	40.1
Quitting contemplation, %		
Not thinking about quitting	40.1	42.3
Thinking about quitting, but no plans to quit	37.1	29.0
Plans to quit in next 6 months	22.7	28.7

Using the exploratory subsample of daily smokers (N = 3,021) and the full sample of nondaily smokers (N = 1,183), we examined the underlying factor structures of the 26-item sets with the software IRTPRO (Cai, du Toit, & Thissen, 2011). The proportion of data missing not by design was negligible (<1%); however, the block administration resulted in substantial planned item-level data missing at random. All analyses used full information maximum likelihood estimation to accommodate the missing data. Local dependence (LD) diagnostic indices (Chen & Thissen, 1997) and high-dimensional exploratory item factor analyses (Cai, 2010) were used to identify clusters of related items, or LD departures from unidimensionality. Item bifactor models (Cai, Yang, & Hansen, 2011; Gibbons & Hedecker, 1992) were then specified to account for these LD clusters.

Examining model results for each smoker type, study team members evaluated items within each specific factor in order to select subsets of items that would collectively be more unidimensional than the initial sets of 26 items. We considered each item's loading on the health expectancies factor, the percentage of common variance accounted for by the health expectancies factor (i.e., item explained common variance or I-ECV; Stucky, Thissen, & Edelen, 2013), and substantive content. Small numbers of items were selected from each item cluster (or specific factor). The two resulting item subsets for daily and nondaily smokers were selected to more closely conform to the unidimensional structure assumed in the final IRT models.

After selecting items for inclusion and removal in this way, the dimensionality of the two resultant item sets was re-evaluated by testing the fit of a one-factor model using the Mplus software (L. K. Muthén & B. O. Muthén, 1998–2010) with weighted least squares mean- and variance-adjusted (WLSMV) estimation for categorical response items and standard model fit indices and criteria (root mean squared error of approximation [RMSEA] ≤ 0.08, Tucker-Lewis index [TLI] ≥ 0.95, comparative fit index [CFI] ≥ 0.95; Browne & Cudeck, 1993; Hu & Bentler, 1999). For daily smokers, model fit was assessed first in the exploratory subsample (N = 3,021) and then confirmed using the confirmatory subsample (N = 1,180); the analysis for nondaily smokers used the full nondaily sample (N = 1,183).

Differential Item Functioning

After identifying and confirming two sufficiently unidimensional item sets to represent health expectancies, the item sets were further evaluated for differential item functioning (DIF). These evaluations were conducted using the full daily (N = 4,201) and nondaily (N = 1,183) smoker samples with IRTPRO (Cai et al., 2011). DIF was evaluated for significance according to gender, race/ethnicity (White, Black, Hispanic), and age (18–30, 31–50, 51+) using established procedures (Edelen, Thissen, Teresi, Kleinman, & Ocepek-Welikson, 2006; Orlando & Marshall, 2002). Items with significant DIF were further evaluated for "impact" by considering the *w*eighted *a*rea *b*etween the expected score *c*urves (wABC) and the expected *d*ifference in *e*xpected *a p*osteriori score (dEAP), indices described in more detail in Hansen et al. Items with wABC values greater than 0.30 were screened for potential removal by evaluating graphical illustrations of the subgroups' expected scores curves, along with the values of the wABC and dEAP indices. Items judged to have non-ignorable DIF were removed from further consideration in their respective item banks (i.e., daily or nondaily).

Calibration of Item Banks

The Health Expectancies item banks for daily and nondaily smokers were concurrently calibrated using data from the full combined sample (N = 5,384, $N_{(\text{daily})}$ = 4,201, $N_{(\text{nondaily})}$ = 1,183). We estimated a two-group IRT model with groups distinguishing daily and nondaily smokers. This calibration, which specified the daily smokers as the reference group, fixed the daily health expectancies mean to 0 and the *SD* to 1 and estimated unique nondaily mean and *SD*. Following PROMIS standards, IRT scores were subsequently rescaled using the *T*-score metric to have a mean of 50 and a *SD* of 10 for daily smokers. The scale for the daily and nondaily group difference was set based on preidentified anchor items whose parameter estimates were constrained to be equal across the groups. Item parameters for non-anchor items were estimated separately for the two groups (see Hansen et al. for more details). The utility of the item banks was determined using IRT-based test information, score precision, and marginal reliability (MR).

Short Form Development

Item parameters from the final calibration were used in the development of a health expectancies fixed-item short form (SF). In order to simplify the administration and scoring of this form, only those items with equal parameters for daily and nondaily smokers (i.e., anchor items in the two-group calibration) were considered for SF inclusion. Among all the possible combinations of eligible items, candidate SFs were identified using selection criteria related to overall content balance, inclusion of items favored by the study team, and the reliability of score estimates across a broad range of the health expectancies continuum. Following PROMIS procedures, SF scoring was based on a transformation of the sum of responses to SF items. The use of summed scores has the particular advantage of allowing for the creation of translation tables by which researchers may convert an observed sum into an IRT-scaled score (Thissen, Nelson, Rosa, & McLeod, 2001). The SF was evaluated using simulated data. For both the daily and nondaily item banks, we examined the reliability of the SF and obtained correlations of SF scores with scores based on the patterns of responses to the full sets of items.

CAT Simulation

Computerized adaptive tests utilize item selection algorithms to administer items that are tailored to the respondent's estimated standing on the measured construct, often resulting in reductions in test length and respondent burden. We conducted CAT simulations using Firestar (Choi, 2009) to evaluate the utility of computer adaptive administration of the daily and nondaily smoker Health Expectancies item banks. These simulations (a) provide an indication of the average number of items from the Health Expectancies item banks that would be administered under typical CAT conditions, (b) indicate which items would be most routinely selected for CAT administration, and (c) characterize the expected CAT-based score reliability.

RESULTS

Item Factor Analyses

Bifactor models, each with four specific factors, were selected to characterize the structure of both the 26 daily smoker items (using the exploratory daily smoker sample) and the 26 nondaily smoker items. In both cases, these models were selected based on their interpretability, comparisons of fit indices, and LD chi-squares. The specific factors identified in the bifactor model represent the content "clusters" in the health expectancies item sets. For the daily smokers, the first specific factor captured common content among 13 items related to physical health effects of smoking (e.g., "Smoking makes me short of breath"; "Smoking makes my lungs hurt"). The second specific factor contained an item pair with similar content (i.e., "smoking is taking years off my life"; "smoking is hazardous to my health"). The third specific factor contained five items expressing worry about future health problems (e.g., "smoking makes me worry about getting cancer"). Finally, the fourth specific factor captured the shared variance in the six items about quitting (e.g., "If I quit smoking I will live longer"). Content clusters for nondaily smokers were similar.

The study team reviewed the bifactor model results for all 26 daily smoker and 26 nondaily smoker items and selected at least one item per specific factor to retain for further consideration in the item banks. Item selection was based on the strength of the general factor loading and item content. For some specific factor item clusters, the I-ECV indicated a strongly unidimensional item loading on the general factor. In these cases, additional items per specific factor were selected.

This process led to the selection of 20 daily smoker items and 18 nondaily smoker items that balanced item content and closely represented the health expectancies dimension. Next, one-factor models were fit to the selected item sets to confirm that they were sufficiently unidimensional. Relative to the original 26 daily smoker items (CFI = 0.934, TLI = 0.928, RMSEA = 0.073), the reduced set of 20 daily smoker items showed improved fit in both the exploratory and confirmatory subsamples (exploratory: CFI = 0.950, TLI = 0.945, RMSEA = 0.082; confirmatory: CFI = 0.962, TLI = 0.958, RMSEA = 0.072) with only a trivial reduction in reliability (MR went from 0.97 to 0.96). Furthermore, in the exploratory subsample, the test-level ECV (Reise, 2012) associated with the health expectancies (general) factor increased substantially from 0.74 to 0.76 indicating a more strongly unidimensional model. Fit indices for the nondaily smokers also suggest a strongly unidimensional 18-item set (CFI = 0.952, TLI = 0.946, RMSEA = 0.079), with improvement in fit compared to the 26-item set (CFI = 0.937, TLI = 0.931, RMSEA = 0.067) and no loss in precision (MR remained 0.96). Similar to daily smoker results, the ECV associated with the health expectancies (general) factor in the nondaily sample solution increased from 0.80 to 0.84.

Differential Item Functioning

Next, the 20 daily and 18 nondaily smoker items underwent DIF testing according to gender, race/ethnicity (White, Black, Hispanic), and age (18–30, 31–50, 51+). For the daily smokers, across all comparisons, two items met the wABC criterion for consideration of removal (i.e., wABC > 0.30), and one item was ultimately removed because of DIF ("Smoking gives me a morning cough", Black vs. White comparison wABC = 0.46). For the nondaily smokers, three items were considered for removal but ultimately retained following review of the expected score curves and impact indices.

Calibration of Item Banks

Using the two-group IRT model with daily smokers as the reference group, 24 total items were calibrated. Within this set, 12 were anchor items (identical item parameters for daily and nondaily smokers) and 1 item had unique item parameters for daily and nondaily smokers. In addition, there were six items in the daily bank and five items in the nondaily bank that were non-overlapping (i.e., items that only occur for that particular smoker group). This process resulted in two Health Expectancies item banks (one for daily and one for nondaily smokers) with a total of 19 and 18 items, respectively. As can be seen in Table 2, the final items tended to be strongly related to the underlying health expectancies construct (*a* parameters for items in both banks ranged from 1.66 to 3.03) and covered a wide range of the health expectancies continuum (*b* parameters ranged from −1.90 to 2.57) that is fairly symmetric around the health expectancies mean.

Figure 1 illustrates the score reliability for the daily and nondaily smoker Health Expectancies item banks (and SF) on a standard *T*-score scale. Full bank scores have reliability values greater than 0.80 from nearly 3 *SD*s below the mean to nearly 3 *SD*s above the mean (i.e., from 20 to 80, in the *T*-score scale). Nondaily smokers had a mean value of 47.5, 0.25 *SD*s below the daily smoker mean of 50. In addition, the nondaily smoker sample had the same amount of health expectancies variability (*SD* = 10) as daily smokers (*SD* = 10).

Health Expectancies Short Form

Evaluation of candidate item sets indicated that six items would provide adequate content coverage and reliability across the health expectancies continuum. The items comprising the six-item SF are indicated in Table 2, and the summed score to IRT score translation table for the SF is contained in Table 3. Figure 1 shows the reduction in score reliability when going from the complete item banks (of 19 and 18 items) to the SF. Despite this reduction, the MR of the SF scores remains quite good (0.87). In addition, these scores correlate strongly (0.95 for daily, 0.96 for nondaily) with those obtained from the complete banks. The results suggest that the six-item SF provides an efficient and reliable measure of health expectancies.

CAT Simulations

CAT simulations were conducted on the daily and nondaily smoker Health Expectancies item banks. Table 2 shows the rate of administration for each item in both daily and nondaily CAT simulations under the stopping rule of a 10-item maximum. As can be seen from these numbers, the rates of item administration are fairly similar across the two smoker types for the common items, and the distribution of administration rates is strongly bimodal for both smoker types; with only a few exceptions, items were either highly likely (e.g., rates from 0.79 to 1.00) or highly unlikely (e.g., rates from 0.00 to 0.19) to be administered. Notably, the items with the higher CAT administration rates do not correspond exactly with those that were selected for the health expectancies SF. This is partly due to the fact that the SF items were selected from among items common to both smoker types, and also because the SF selection considered breadth of content in addition to empirical information. Table 4 provides the results of simulations that used a *SE* of 3.0 (in the *T*-score metric) as the CAT stopping criterion, which corresponds to a reliability of slightly greater than 0.90, and a range of limits on

Table 2. Health Expectancies Item Banks for Daily and Nondaily Smokers

		CAT		Item parameters				
Item[a]	D/ND	D	ND	a	b_1	b_2	b_3	b_4
I worry that smoking will lower my quality of life (SF)	Both	0.90	1.00	3.03	−0.89	−0.15	−0.47	−0.96
Smoking makes me worry about getting heart troubles (SF)	Both	0.90	1.00	2.96	−0.93	−0.10	−0.52	1.06
Smoking is taking years off my life (SF)	Both	0.30	0.94	2.64	−1.46	−0.49	−0.26	−0.80
Smoking makes me short of breath (SF)	Both	0.15	0.19	1.94	−1.01	−0.10	−0.84	1.51
Smoking causes me to get tired easily (SF)	Both	0.11	0.04	1.78	−0.15	−0.60	1.44	2.07
Smoking irritates my mouth and throat (SF)	Both	0.05	0.02	1.58	−0.34	−0.85	1.81	2.57
Smoking makes me worry about getting emphysema	Both	0.79	0.98	2.91	−1.01	−0.19	−0.43	−0.94
If I quit smoking I will breathe easier	Both	0.06	0.10	2.48	−1.90	−1.03	−0.27	−0.35
If I quit smoking I will feel more energetic	Both	0.08	0.54	1.95	−1.30	−0.57	−0.44	1.15
Smoking makes my lungs hurt	Both	0.10	0.05	1.79	−0.10	−0.99	1.91	2.53
Smoking causes damage to my gums and teeth	Both	0.03	0.14	1.73	−1.50	−0.41	−0.50	1.18
Smoking leaves an unpleasant taste in my mouth	D[b]	0.07		1.70	−1.02	−0.07	−0.85	1.53
Smoking leaves an unpleasant taste in my mouth	ND[b]		0.03	1.35	−1.77	−0.38	−0.49	1.35
It takes me longer to recover from a cold because I smoke	Both	0.12	0.19	1.66	−0.41	−0.36	1.13	1.82
Smoking makes me worry about getting cancer	D	1.00		3.35	−1.08	−0.30	−0.34	−0.81
If I quit smoking I will be healthier	D	0.40		2.85	−2.11	−1.27	−0.46	−0.11
Smoking is hazardous to my health	D	0.16		2.64	−2.32	−1.31	−0.59	−0.00
If I quit smoking I will live longer	D	0.08		2.37	−1.77	−0.98	−0.06	−0.52
If I quit smoking I will get instant health benefits	D	0.02		1.62	−1.48	−0.53	−0.41	1.13
Smoking gives me a headache	D	0.01		1.09	1.38	2.47	3.74	4.67
Smoking makes me feel weaker physically	ND		0.34	2.51	−0.50	−0.36	1.11	1.95
Smoking makes me worry about getting high blood pressure	ND		0.16	2.23	−0.60	−0.14	−0.80	1.30
Smoking makes it harder for me to exercise or play sports	ND		0.29	2.18	−0.88	−0.05	−0.71	1.30
Smoking gives me a morning cough	ND		0.01	1.41	−0.12	1.01	1.88	2.56
Smoking leaves a stain on my fingers	ND		0.00	1.24	0.11	1.09	1.88	2.69

Note. SF = short form. D/ND column indicates if the item parameters were identical in daily and nondaily groups (both), unique to only the daily group (D), or unique to only the nondaily group (ND). CAT column indicates the rate of item administration for the 10-item maximum condition. Item slope and threshold parameters were obtained through calibrations of the full item banks.
[a]All items used the following response options: 0 = not at all, 1 = a little bit, 2 = somewhat, 3 = quite a bit, 4 = very much.
[b]Items with unique item parameters in both the daily and nondaily groups.

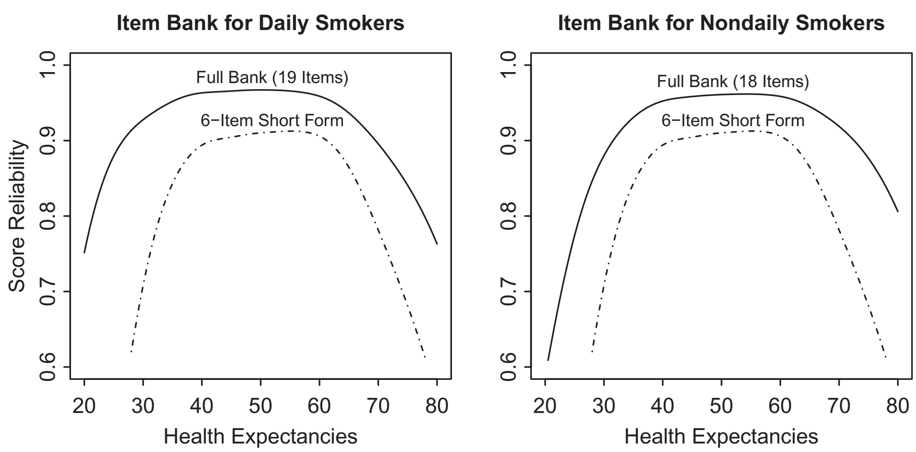

Figure 1. Health Expectancies item bank and short form reliability curves for daily and nondaily smokers.

the maximum number of items allowed to be administered (4, 6, 8, 10, 12). To summarize these results, the correlation between CAT and full bank scores was 0.96 and higher for the conditions examined. For a maximum test length of eight items, CAT terminated with a *SE* of less than 3.0 after administering an average of 5.1 for daily smokers and 5.7 items for nondaily smokers.

DISCUSSION

As part of the PROMIS Smoking Initiative, we developed comprehensive item banks to assess the negative health expectancies of smoking among daily and nondaily smokers. The item banks include a total of 24 items; 13 items are common across smoker types, 6 are unique to daily smokers, and 5 are unique to nondaily smokers. The psychometric properties of the Health Expectancies item banks are excellent; they are strongly unidimensional, highly reliable (reliability = 0.95 and 0.96 for daily and nondaily smokers, respectively), and perform similarly across gender, age, and race/ethnicity groups.

The item banks were developed to provide researchers and practitioners with a flexible system for assessing the negative health expectancies of smoking. In addition to the full item banks, we developed a six-item SF that is applicable across daily and nondaily smokers. This SF demonstrates high reliability (MR = 0.87) and its content represents a broad range of the health expectancies continuum. Researchers may also choose to administer items from the Health Expectancies item banks adaptively. Our CAT simulation results showed that the health expectancies construct can be assessed adaptively with excellent precision with an average of 5–6 items. The Health Expectancies item banks and SF will be available for public use through inclusion in the larger PROMIS library, and a free online tool for administering adaptive tests is available through the PROMIS Assessment Center (www.assessmentcenter.net). The smoking assessment toolkit products are also available for download from the project Web site (http://www.rand.org/health/projects/promis-smoking-initiative.html).

Our approach was designed to encourage uptake of the banks by researchers and practitioners. First, we used existing measures to form the basis of the item banks' content. This both increases researchers' familiarity with much of the item content as well as facilitates future work to establish the comparability of results from these banks to those from studies using similar items from existing inventories. Despite our use of existing measures as a basis, our inclusion of information from focus group discussions with smokers as well as our use of state-of-the-art psychometric methods to develop the Health Expectancies item banks represents a considerable improvement beyond current measurement efforts. The assessment system is based on well-established content, has been updated with input from current smokers, and has been formulated using IRT, which results in a robust, versatile, and sustainable framework for measurement. The item banks provide a variety of psychometrically strong assessment options from the preselected SF, to tailored SFs, to CATs, all of which serve to minimize respondent burden considerably and afford researchers with more time and space to assess other constructs of interest. Further, as our understanding of the negative health expectancies of smoking changes over time, so can the content breadth of the item banks and their resultant assessment forms without the need to distribute new "versions" as is typical of existing questionnaires.

Because our developmental process started by considering existing item content in the literature, there is very little "new" item content in the Health Expectancies banks. However, it is interesting to note that the existing instruments from which these banks' content draws are not health-specific (e.g., PRBQ, McKee et al., 2005; SEQ, Rohsenow et al., 2003). In fact, despite the clear importance of health expectancies in decisions to start, maintain, and attempt to quit smoking (McKee et al, 2005; Romer & Jamieson, 2001), there are currently no

Table 3. Health Expectancies Summed Score to Scaled Score Translation Table for the Six-Item SF

Six-item short form		
Summed score	Scaled score (T)	*SE*
0	30.6	5.2
1	35.4	4.1
2	38.2	3.7
3	40.5	3.5
4	42.5	3.4
5	44.2	3.4
6	45.7	3.3
7	47.2	3.3
8	48.5	3.2
9	49.8	3.2
10	51.1	3.2
11	52.3	3.2
12	53.5	3.2
13	54.8	3.2
14	56.0	3.2
15	57.2	3.3
16	58.5	3.4
17	59.9	3.5
18	61.3	3.6
19	62.8	3.8
20	64.4	4.0
21	66.1	4.2
22	68.1	4.5
23	70.2	4.7
24	73.5	5.4

Table 4. Simulated Adaptive Tests for the Health Expectancies Item Banks

	Maximum no. of items					
	4	6	8	10	12	All items
Daily smokers						
Average items administered	3.93	4.79	5.12	5.33	5.48	19
Proportion receiving maximum items	0.93	0.30	0.15	0.10	0.07	1
Marginal reliability	0.88	0.90	0.90	0.91	0.91	0.96
r (T_{CAT}, T_{full})	.96	.97	.98	.98	.98	1.00
Nondaily smokers						
Average items administered	4.00	5.21	5.71	6.01	6.22	18
Proportion receiving maximum items	1.00	0.46	0.21	0.13	0.10	1
Marginal reliability	0.86	0.89	0.90	0.90	0.90	0.95
r (T_{CAT}, T_{full})	.96	.98	.98	.98	.98	1.00

measures developed specifically to assess health expectancies of smoking. It is hoped that these new item banks will be particularly useful in helping identify individuals who are ready to quit smoking and might respond to a brief intervention by a primary care provider (e.g., motivational interview). It is also plausible to envision that measurement of this construct could help fine tune smoking-related intervention strategies that attempt to capitalize on the relationship between health-related outcome expectancies and smoking behavior (Bize et al., 2012; Schlam & Baker, 2013).

Preliminary validity of the health expectancies scores is evaluated in a separate paper in this issue (Edelen, Stucky, et al.). This evaluation, which includes associations of health expectancies scores with scores on the other PROMIS smoking banks, scores on other measures of quality of life, and patterns of mean scores according to demographic and smoking characteristics, has yielded promising results. For example, analyses have shown that higher health expectancies scores are associated with higher interest in quitting and higher likelihood of having a recent quit attempt. Our team is currently collecting data from a subset of the original calibration sample used in this paper as well as from a community-based sample of smokers to support further validity evaluation. Specifically, we plan to use the new data to evaluate test—retest reliability; stability of the health expectancies construct over time; associations of health expectancies scores with health care utilization, use of other tobacco products, quitting history and future quitting plans; and feasibility of cross-walking scores from these banks to commonly used measures.

In addition to these ongoing efforts, other areas of interest for future research include a more controlled examination of the relationship between health expectancies scores and cessation. For example, it would be useful to determine the extent to which realization of health expectancies may indicate readiness to quit. Due to the versatility of the available assessment options for our Health Expectancies item banks, future research could also facilitate regular assessment of health expectancies in primary care settings, thus enabling identification of opportunities for effective brief intervention as mentioned above. More generally, the Health Expectancies item banks promise to provide researchers and practitioners with an efficient and flexible measurement tool for the reliable, valid, and standardized assessment of this key smoking-related construct.

FUNDING

This work was supported by the National Institute on Drug Abuse (R01DA026943 to MOE). Additionally, MH was supported by the Institute for Education Sciences with a predoctoral training grant (R305B080016), and LC was supported by the National Institute on Drug Abuse (R01DA030466).

DECLARATION OF INTERESTS

None declared.

ACKNOWLEDGMENTS

The authors would like to thank the PROMIS Smoking Initiative Advisory Group: Ronald D. Hays and Michael Ong, UCLA; David Cella, Feinberg School of Medicine, Northwestern University; Daniel McCaffrey, Educational Testing Service; Raymond Niaura, American Legacy Foundation, Brown University; Paul Pilkonis, University of Pittsburgh; and David Thissen, University of North Carolina at Chapel Hill.

REFERENCES

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall. ISBN 0-13-815614-X

Bize, R., Burnand, B., Mueller, Y., Rège-Walther, M., Camain, J. Y., & Cornuz, J. (2012). Biomedical risk assessment as an aid for smoking cessation. *Cochrane Database of Systematic Reviews*, *12*, CD004705. doi:10.1002/14651858.CD004705.pub4

Brandon, T. H., & Baker, T. B. (1991). The Smoking Consequences Questionnaire: The subjective expected utility of smoking in college students. *Psychological Assessment*, *3*, 484–491.

Brandon, T. H., Herzog, T. A., Irvin, J. E., & Gwaltney, C. J. (2004). Cognitive and social learning models of drug dependence: Implications for the assessment of tobacco dependence in adolescents. *Addiction (Abingdon, England)*, *99*(Suppl. 1), 51–77. doi:10.1111/j.1360-0443.2004.00737.x

Brewer, N., Chapman, G., Gibbons, F., Gerard, M., McCaul, K., & Weinstein, N. (2007). Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination. *Health Psychology*, *26*, 136–145. doi:10.1037/0278-6133.26.2.136

Brown, R. (2003). Intensive behavioral treatment. In D. B. Abrams, R. Niaura, R. Brown, K. Emmons, M. G. Goldstein, & P. M. Monti (Eds.), *The tobacco dependence treatment handbook* (pp. 178–229). New York, NY: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Cai, L. (2010). High-dimensional exploratory item factor analysis by a Metropolis-Hastings Robbins-Monro algorithm. *Psychometrika*, *75*, 33–57. doi:10.1007/s11336-009-9136-x

Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software]*. Chicago, IL: Scientific Software International.

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, *16*, 221–248. doi:10.1037/a0023350

Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., … Rose, M. (2007). The patient-reported outcomes measurement information system (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Medical Care*, *45*, S3–S11. doi:10.1097/01.mlr.0000258615.42478.55

Chen, W.-H., & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*, 265–289. doi:10.2307/1165285

Choi, S. W. (2009). Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*, *33*, 644–645. doi:10.1177/0146621608329892

Copeland, A. L., Brandon, T. H., & Quinn, E. P. (1995). The Smoking Consequences Questionnaire-Adult: Measurement of smoking outcome expectancies of experienced smokers. *Psychological Assessment*, *7*, 484–494. doi:10.1037/1040-3590.7.4.484

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

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research*, *16*(Suppl. 1), 5–18. Retrieved from http://rd.springer.com/journal/11136

Edelen, M. O., Stucky, B. D., Hansen, M., Tucker, J. S., Shadel, W. G., & Cai, L. (2014) The PROMIS® smoking initiative: initial validity evidence for six new smoking item banks. *Nicotine & Tobacco Research, 16*, S249–S259.

Edelen, M. O., Thissen, D., Teresi, J., Kleinman, M., & Ocepek-Welikson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: Application to the mini-mental status examination. *Medical Care*, *44*(Suppl. 3), S134–S142. doi:10.1097/01.mlr.0000245251.83359.8c

Edelen, M. O., Tucker, J. S., Shadel, W. G., Stucky, B. D., & Cai, L. (2012). Toward a more systematic assessment of smoking: Development of a smoking module for PROMIS. *Addictive Behaviors*, *37*, 1278–1284. doi:10.1016/j.addbeh.2012.06.016

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gibbons, R. D., & Hedeker, D. (1992). Full-information item bifactor analysis. *Psychometrika*, *57*, 423–436. doi:10.1007/BF02295430

Hansen, M., Cai, L., Stucky, B. D., Tucker, J. S., Shadel, W. G., & Edelen, M. O. (2014) Methodology for developing and evaluating the PROMIS® smoking item banks. *Nicotine and Tobacco Research*, *16*, S174–S188.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. doi:10.1080/10705519909540118

McKee, S. A., O'Malley, S. S., Salovey, P., Krishnan-Sarin, S., & Mazure, C. M. (2005). Perceived risks and benefits of smoking cessation: Gender-specific predictors of motivation and treatment outcome. *Addictive Behaviors*, *30*, 423–435. doi:10.1016/j.addbeh.2004.05.027

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Orlando, M., & Marshall, G. (2002). Differential item functioning in a Spanish translation of the PTSD Checklist: Detection and evaluation of impact. *Psychological Assessment*, *14*, 50–59. doi:10.1037/1040-3590.14.1.50

Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., … Cella, D. (2007). Psychometric evaluation and calibration of health-related quality of life items banks: Plans for the patient-reported outcome measurement information system (PROMIS). *Medical Care*, *45*(5 Suppl. 1), S22–S31. doi:10.1097/01.mlr.0000250483.85507.04

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*, 667–696. doi:10.1080/00273171.2012.715555

Rohsenow, D., Abrams, D., Monti, P., Colby, S., Martin, R., & Niaura, R. (2003). The Smoking Effects Questionnaire for adult populations. Development and psychometric properties. *Addictive Behaviors*, *28*, 1257–1270. doi:10.1016/S0306-4603(02)00254-X

Romer, D. & Jamieson, P. (2001). The role of perceived risk in starting and stopping smoking. In P. Slovic (Ed.), *Smoking: Risk, perception, & policy* (pp. 64–80). Thousand Oaks, CA: Sage Publications.

Schlam, T. R., & Baker, T. B. (2013). Interventions for tobacco smoking. *Annual Review of Clinical Psychology*, *9*, 675–702. doi:10.1146/annurev-clinpsy-050212-185602

Stead, L. F., Buitrago, D., Preciado, N., Sanchez, G., Hartmann-Boyce, J., & Lancaster, T. (2013). Physician advice for smoking cessation. *Cochrane Database of Systematic Reviews*, *5*. Art. No.: CD000165. doi:10.1002/14651858.CD000165.pub4

Strasser, A., Tang, K., Romer, D., Jepson, C., & Cappella, J. (2012). Graphic warning labels in cigarette advertisements: Recall and viewing patterns. *American Journal of Preventive Medicine*, *43*, 41–47. doi:10.1016/j.amepre.2012.02.026

Stucky, B. D., Thissen, D., & Edelen, M. O. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement*, *37*, 41–57. doi:10.1177/0146621612462759

Thissen, D., Nelson, L., Rosa, K., & McLeod, L. D. (2001). Item response theory for items scored in more than two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 141–186). Mahwah, NJ: Lawrence Erlbaum & Associates.

U.S. Department of Health and Human Services. (2004) *The health consequences of smoking: A report of the surgeon general*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. Retrieved January 10, 2012, from http://www.cdc.gov/tobacco/data_statistics/sgr/2004/index.htm

Weinstein, N. D. (1993). Testing four competing theories of health-protective behavior. *Health Psychology*, *12*, 324–333. doi:10.1037/0278-6133.12.4.324

Weinstein, N. D. (2001). Smokers' recognition of their own vulnerability to harm. In P. Slovic (Ed.), *Smoking: Risk, perception, & policy* (pp. 81–96). Thousand Oaks, CA: Sage Publications.

Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Education and Behavior*, *27*, 591–615. doi:10.1177/109019810002700506