



Published in final edited form as:

*Respir Med.* 2014 February ; 108(2): 252–263. doi:10.1016/j.rmed.2013.12.008.

## Developing an item bank and short forms that assess the impact of asthma on quality of life\*

Brian D. Stucky<sup>a,\*</sup>, Maria Orlando Edelen<sup>b</sup>, Cathy D. Sherbourne<sup>a</sup>, Nicole K. Eberhart<sup>a</sup>, and Marielena Lara<sup>a</sup>

Brian D. Stucky: bstucky@rand.org

<sup>a</sup>RAND Corporation, Santa Monica, CA 90407, USA

<sup>b</sup>RAND Corporation, Boston, MA 02116, USA

### Summary

The present work describes the process of developing an item bank and short forms that measure the impact of asthma on quality of life (QoL) that avoids confounding QoL with asthma symptomatology and functional impairment. Using a diverse national sample of adults with asthma ( $N = 2032$ ) we conducted exploratory and confirmatory factor analyses, and item response theory and differential item functioning analyses to develop a 65-item unidimensional item bank and separate short form assessments. A psychometric evaluation of the RAND Impact of Asthma on QoL item bank (RAND-IAQL) suggests that though the concept of asthma impact on QoL is multi-faceted, it may be measured as a single underlying construct. The performance of the bank was then evaluated with a real-data simulated computer adaptive test. From the RAND-IAQL item bank we then developed two short forms consisting of 4 and 12 items (reliability = 0.86 and 0.93, respectively). A real-data simulated computer adaptive test suggests that as few as 4–5 items from the bank are needed to obtain highly precise scores. Preliminary validity results indicate that the RAND-IAQL measures distinguish between levels of asthma control. To measure the impact of asthma on QoL, users of these items may choose from two highly reliable short forms, computer adaptive test administration, or content-specific subsets of items from the bank tailored to their specific needs.

### Keywords

Item response theory; Psychometrics; Health-related; quality of life; Asthma; RAND-IAQL; Item bank

---

\*Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under Award Number R01HL10732. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

© 2013 Elsevier Ltd. All rights reserved.

\*Corresponding author. Tel.: +1 310 393 0411x6336.

**Conflict of interest:** I (we) certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

## Introduction

According to the U.S. National Heart, Lung, and Blood Institute's (NHLBI) Asthma Guidelines, the goal of asthma treatment is to improve the quality of life (QoL) of people who have asthma, while working toward controlling symptoms, reducing the risk of exacerbations, and preventing asthma-related death [1].

Recently, leaders in the asthma field noted important limitations of existing asthma-specific QoL measures [2]. Most notably, past efforts to measure asthma QoL have resulted in tools that can confound QoL with symptoms (e.g., shortness of breath, wheezing), functional impairment (e.g., limitations in daily activities), and control (i.e., the extent to which symptoms, functional impairments, and risks of negative events are minimized and goals for treatment are met). The majority of these assessments also lack the patient's perception of the *impact* or *bother* of asthma symptoms on his or her life.

In light of these limitations, the Asthma Quality of Life Subcommittee of the 2010 NHLBI Asthma Outcomes Workshop declined to recommend any existing instrument as a core outcome measure of asthma-specific QoL [1,2]. Instead, the Subcommittee strongly recommended development of new instruments that incorporate the patient's perspective and are able to measure the impact of asthma on QoL as a construct that is distinct from asthma symptoms or functional status. The primary objective of the present work responds to this recommendation by developing new freely available instrumentation for measuring the impact of asthma on QoL that avoids confounding QoL with asthma symptomatology and functional impairment, and includes many domains of life important to people with asthma.

Our developmental process began with formative work, a detailed description of which can be found in Eberhart et al. [3]. Briefly, although the development of our item pool incorporated literature review and expert recommendations, the majority of its content was generated based on feedback from adults with asthma who participated in focus groups. Salient themes generated from focus group discussions included both general (e.g., enjoyment of life) and specific (e.g., sleep difficulty, affect, medication, physical activities, social relations, health) areas of impact. Using the focus group transcripts, we followed a well-defined item development and refinement process, to arrive at a set of items in standard format representing a wide range of content regarding the impact of asthma on QoL.

This paper describes the development and psychometric properties of an item bank to measure the impact of asthma on QoL in adults. Using data from a large national field test of adults with asthma, we evaluated the pool of candidate items using modern psychometric methods, including item response theory (IRT) and computerized adaptive tests. Our analytic plan adheres to many guidelines used by the patient reported outcomes measurement information system (PROMIS) collaborative [4]. Following these guidelines the graded response model (GRM [5]), is used to “calibrate” (or characterize) the strength of the relationship between items and the construct being measured (here the impact of asthma on QoL) and the location on the construct's scale where the item is most informative.

The collection of calibrated items is referred to as an “item bank.” Item banks – large sets of items that each measures the same underlying construct – have many advantages over

traditional scales. Because not all the items in the bank need to be administered in order to produce a reliable score, item banks provide a very flexible assessment environment. For example, one of the unique features of item banks is that items can be administered adaptively (i.e., with computer adaptive testing), often resulting in reduced overall test lengths. However, for situations in which it is impractical to administer a computer adaptive test, reliable subsets of items can be drawn from the bank to produce traditional, brief fixed-length instruments (i.e., short forms) that can be administered via computer or paper and pencil. Items may be selected for short forms to achieve various measurement goals. For example, if the goal is to assess the impact of asthma on QoL among a non-clinical sample of people with a wide range of potential asthma impact, one would select items that optimize measurement precision across the entire impact continuum. Alternatively, a study involving treatment-seeking patients with severe asthma may benefit most from items that provide precision at the higher end of the asthma impact continuum, whereas an intervention study aimed at improving the social QoL for people with asthma might select a short form that over-represents content specific to that goal (e.g., items that assess the impact of asthma on social activities).

The present work develops an item bank and separate short forms that measure the impact of asthma on QoL. Our short forms were selected to be brief and representative with respect to breadth of content while maintaining acceptable measurement precision across the entire impact continuum. Other short forms that focus on a particular QoL component (e.g., impact of asthma on social concerns) can be generated from the item bank if desired.

## Methods

### Participants and procedures

A national sample of adults (ages 18+) with asthma ( $N = 2032$ ) was recruited by Harris Interactive, a global interactive media and services company, and all survey measures were completed via internet assessment. All study procedures were approved by the institution's IRB. Participants were eligible for the study if (1) they had been told by a doctor or other health professional that they had asthma, and (2) they reported still having asthma. To assure that we would have variability across a range of asthma severity, we also required that 90% of the sample had experienced an episode of asthma or an asthma attack during the prior 12 months [6]. We sampled Hispanic, Black, Asian and non-Hispanic Whites, over-sampling minorities to have at least 200 participants in each group. Similarly, we targeted at least 200 participants within each of four age groups (18–34, 35–49, 50–64 and 65+), and achieved a distribution of about 40% men to reflect the distribution of individuals with asthma in the general U.S. adult population. Because of concerns about confounding Chronic Obstructive Pulmonary Disease (COPD) with asthma, we limited the proportion of the sample that had comorbid COPD. As incentives, participants received points through Harris that can be redeemed for rewards such as an Amazon gift card.

Table 1 describes the demographic characteristics and health service utilization patterns of participants. In the past year, 39% ( $N = 785$ ) of the sample had visited an emergency room or urgent care facility, and 19% had spent at least one night in a hospital because of asthma.

A subset of the sample reported having chronic medical conditions; 10% ( $N = 194$ ) of the sample had chronic heart disease and 14% ( $N = 287$ ) had COPD.

## Measures

**Impact of asthma on QoL item pool**—Our pool of candidate items consisted of 112 items measuring various aspects of the impact of asthma on QoL. These items were developed using focus groups, literature review, expert input, and cognitive interviews (see Eberhart et al. [3], for a detailed description of the item development process). In cases where items were selected from the literature review, permission to use such items was sought and items were incorporated into the preliminary item bank only if permission was granted [7–15]. Items were standardized to have a consistent timeframe (past 4-weeks), orientation (first-person), and response format (5-point Likert-type) reflecting magnitude (i.e., “not at all” to “very much”) or frequency of impact (i.e., “never” to “almost always”). The order of item administration was randomized to avoid serial effects [16].

**Asthma control test**—The Asthma Control Test (ACT) [17] is a five-item measure that includes content on asthma symptoms, use of rescue medication, impact on functioning, and a self-rating of asthma control. Each item uses a 5-point Likert response scale. The total score ranges from 5 (poor control) to 25 (good control) and the validated score categories are 5–15 (poorly controlled), 16–19 (somewhat controlled), and 20–25 (well controlled) [18].

**Additional information**—We collected information on demographics, asthma symptoms, co-morbid health conditions (e.g., COPD, sinusitis, etc), and asthma-related health care use (e.g., emergency department, hospitalization).

## Analytic approach

**Factor analysis and item reduction**—We began by randomly splitting our total sample into exploratory ( $N = 1500$ ) and confirmatory ( $N = 532$ ) sub-samples. Analyses initially used only the exploratory sub-sample; the confirmatory subsample was set aside to be used as an independent check on the validity of the dimensionality findings. The goal of the factor analyses was to identify unidimensional sets of items that could form the basis for the item bank(s). All factor analyses were conducted using the computer program *Mplus* [19] and the mean and variance adjusted weighted least squares algorithm (WLSMV) that is appropriate for categorical response items. Model fit was evaluated with commonly used model fit indices (RMSEA = 0.08, TLI = 0.95, CFI = 0.95) [9,20].

**Exploratory factor analysis (EFA)**—Given that preliminary qualitative item development work generated a number of distinct content categories (see Eberhart et al. [3]), we anticipated that multiple dimensions might be needed to explain item responses. However, through the course of considering several multiple EFA solutions the model fit evidence overwhelmingly supported a single-factor solution (these analyses are described in more detail in the results section). Thus analyses proceeded assuming a single latent dimension for the item pool.

**Confirmatory factor analysis (CFA)**—Following EFA, we used modification indices from 1-factor CFA models to identify clusters or pairs of items with excess dependence. The presence of local dependence violates the IRT assumption of unidimensionality and could result in misleading score estimates; therefore it is necessary to identify and minimize local dependence in the item bank. Thus, for a given pair of locally dependent items or group of items, we considered the items' factor loadings and wording to determine which item(s) to remove. This model fitting process was repeated iteratively until no additional local dependence was identified, at which point a final 1-factor CFA model was fit to the confirmatory sample to cross-validate dimensionality findings.

**Differential item functioning**—Once a provisional set of unidimensional items was identified, the complete sample ( $N = 2032$ ) was used to test for differential item functioning (DIF) using item response theory (IRT) within the computer program IRTPRO [21]. Item-level DIF indicates that responses to an item for members of subgroups vary in a way that is not predicted by the IRT model after accounting for group-level mean and variances differences. Thus, the IRT model does not hold and the item should be considered for removal. DIF was tested according to age groups, education, race/ethnicity, and gender.

Analysis of DIF used three steps. First, two-group chi-square tests were evaluated across subgroups and the combined significance tests for all comparisons per grouping variable were adjusted using the Benjamini-Hochberg procedure [22] at  $p < 0.05$ . Next, to evaluate the magnitude of DIF, items demonstrating significant DIF after  $p$ -value correction were further probed by computing the weighted area between the expected score curves based on an approach by Raju (1988) [23]. Our experience indicates that using corrected significance tests when coupled with effect size indicators is useful in revealing the items most likely to result in bias across subgroups. Finally, for items that met both criteria we examined plots of expected item scores generated from the parameters of the DIF model that best fit the data prior to selecting items for removal.

### Item calibration

To characterize the final set of items as an item bank we calibrated the data using the graded IRT model (GRM) [5]. For each item, the GRM characterizes the relationship of the item responses with the underlying latent construct with a unique *slope* parameter ( $a$ ), and indicates each response option's location along the continuum (of asthma impact) with four threshold parameters ( $b_k$ ) (for five response option items). To ensure that there was adequate power to estimate the threshold parameters, prior to conducting the IRT analysis we evaluated the response coverage across all response options. IRT-based assessments of latent constructs (here the negative impact of asthma on QoL) allow the items' slope or discrimination parameters to vary as a function of the strength of the relationship between the item and the construct, as opposed to Rasch-based IRT techniques that select items having an equivalent relation to the construct so that their slope parameters may be fixed to equality (1). In addition, the IRT model is scaled by assuming a normal underlying distribution ( $N(0,1)$ ), though graphical illustrations of score precision and tables of score estimates reported here follow standard PROMIS conventions and translate the Z-score metric to a T-score scale with a mean of 50 and a standard deviation of 10.

## Computer adaptive test simulations

The performance of the item bank was evaluated using “real-data” computer adaptive test simulations from the full sample of participants. A real-data computer adaptive test assesses the performance of the item bank as if the respondents in the calibration sample had received the items adaptively and stopped when a predetermined level of precision was reached (i.e., a computer adaptive test administration) rather than answering every item in the bank. Computer adaptive test simulations were conducted to (1) evaluate the overall performance of the item bank; (2) provide an expectation of the number of items administered under typical computer adaptive test conditions; (3) indicate the items most routinely administered in computer adaptive test scenarios; and (4) provide a comparison to short form scores. In this simulation, the computer adaptive test was programmed to stop administering items after achieving a score standard error of 0.316 (which corresponds to a reliability of 0.90) or completing administration of 12 items, whichever came first.

## Short form development, scoring, and preliminary validity evidence

Two short forms were created to provide highly reliable fixed-length assessments of the impact of asthma on QoL. The goal for the first short form was to generate a reliable assessment across the range of asthma impact scores using the fewest, most widely relevant items. Thus items were selected that were reflective of the impact of asthma on ‘global’ aspects of QoL and provided information (i.e., according to the IRT model) across the range of impact of asthma on QoL. In order to provide a more content-diverse short form, the second (longer) short form added items from several content domains that Eberhart et al. [3] identified based on a series of focus groups as being of particular interest to adults with asthma (e.g., physical limitations, social concerns (see Eberhart et al. [3])).

Following standard practice (for examples, see Irwin et al. [24]; DeWitt et al. [25]), we used the sum of the item responses to generate IRT-based scores for each short form [26] and rescaled them to a T-score metric with a mean of 50 and a standard deviation of 10. For example, a score of 40 is one standard deviation below the mean and suggests less impact of asthma on QoL.

We compared the precision of the computer adaptive test and short form scores to scores based on the full bank (which in this case represents the gold standard) using root mean square errors (RMSE). The RMSE is the square root of the average error variance (the squared standard error of measurement of the score) across respondents, and indicates the average precision of the IRT score estimates. Finally, to provide an initial indication of validity, scores for the short forms and real-data computer adaptive test were compared against asthma control categories derived from the ACT.

## Results

### Factor analysis and item reduction

**Exploratory factor analysis**—Initial EFA solutions using the 112-item pool and the exploratory sample ( $N = 1500$ ) focused on capturing the relationships among the item responses using high and low-dimensional models. Through the course of considering

multiple factor solutions the model fit evidence overwhelmingly supported a single-factor solution ( $\chi^2 = 37,898$ ,  $df = 6104$ ; CFI = 0.945; TLI = 0.944; RMSEA = 0.059), which accounted for 71% of the total variance explained.

**Confirmatory factor analysis**—Consistent with the goal of producing a unidimensional set of items, we next used the exploratory sample to identify and remove items with local dependence. Using results from a 1-factor CFA model involving all 112 items, we elected to remove 26 items because of local dependence. For example, the item pair: “I had to make compromises because of the cost of treating my asthma” and “The cost of treating my asthma was a burden to me” displayed strong local dependence because of the overlapping content. For this particular item pair, we elected to remove the latter item because it was also locally dependent with several other items. A 1-factor CFA model using the remaining 86 items revealed the presence of additional, though weaker, local dependence from which 14 more items were removed. A final 1-factor CFA of the remaining 72 items did not reveal any additional instances of problematic local dependence, and fit was acceptable in both the exploratory ( $\chi^2 = 15,683$ ,  $df = 2484$ , CFI = 0.965, TLI = 0.964, RMSEA = 0.060) and confirmatory ( $\chi^2 = 6282$ ,  $df = 2484$ ; CFI = 0.971, TLI = 0.970, RMSEA = 0.054) samples.

**Differential item functioning**—DIF was evaluated among the remaining 72 items using the combined sample ( $N = 2032$ ) by gender, age (18–34, 35–54, and 55–64 years), race/ethnicity (African American, Asian, Hispanic, and non-Hispanic White), and educational status (up to completion of high school versus attending some high school and greater). DIF comparisons of educational status did not indicate the need for any item removal. DIF tests according to gender resulted in multiple items with statistically significant DIF. However, closer examination indicated only minor gender DIF impact (wABC's < 0.20) that was not substantial enough to warrant item removal.

Age and race/ethnicity DIF comparisons led to a total of seven items being removed. The item “I worried about becoming addicted to my asthma medication” had significant race/ethnicity DIF between White and Asian subgroups (wABC = 0.31), and between Black and Asian subgroups (wABC = 0.34), and was removed. Six additional items were removed because of age DIF. Fig. 1 contains two example items that were removed because of age DIF (“I was bothered by the unwanted attention I got because of my asthma” and “It was hard for me to admit that I have asthma”). These items illustrate age bias such that at a given level of impact on the T-score metric ( $x$ -axis), higher item responses ( $y$ -axis) are more likely from younger individuals. This means that, given mean and variance group differences, younger individuals are more likely to report being bothered by unwanted attention and not wanting to admit having asthma. For these six DIF items it was typically the younger (18–34) to older (55–64) age group comparison that resulted in significant and problematic DIF with wABC effect sizes ranging from 0.31 to 0.56.

### Final IRT calibration

Our analytic process resulted in a 65-item unidimensional bank of items measuring the Impact of Asthma on QoL (hereafter referred to as the RAND-IAQL). Final IRT item parameters are presented in Table A1; items are sorted by magnitude of the slope parameter.

Fig. 2 displays the item bank's reliability (y-axis) across the impact continuum (x-axis; mean = 50; SD = 10). Unsurprisingly given the number of items in the bank, marginal reliability levels are 0.90 or higher from nearly two standard deviations below the mean to three standard deviations above the mean (item bank marginal reliability = 0.98).

### Computer adaptive test simulation

Computer adaptive test simulation results indicated that relatively few items were needed to obtain a RAND-IAQL score estimate with reliability of 0.90 (mean = 4.97, SD = 3.30). The variability in the precision of the computer adaptive test score estimates (i.e., standard error of the score estimate) ranged from a minimum of 0.25 to a maximum of 0.49 (mean SE = 0.30). The correlation between the computer adaptive test -estimated and full bank-estimated RAND-IAQL scores was high as expected ( $r = 0.96$ ).

Item-by-item exposure information for the computer adaptive test simulation is contained in the right-hand column of Table A1. One item (“I felt like I couldn't enjoy life because of my asthma”) was administered to more than 50% of respondents, an additional four items were administered to between 30 and 50% of respondents, six items were administered to between 15 and 25% of respondents, 23 items were administered at least once, but to fewer than 15% of respondents, and 31 items were not administered in the computer adaptive test simulation we conducted.

The distribution of the number of items administered to each respondent in the computer adaptive test simulation is shown in Fig. 3. Recall that in a computer adaptive test, items are administered until a pre-determined level of score precision is met (in this case, reliability = 0.90) or the maximum number of items is administered (in this case 12), thus the number of items administered to each respondent is reflective of the item bank's quality (the fewer the better). As can be seen in this figure, 25% of respondents (y-axis) were scored reliably based on responses to just two items (x-axis), and over 80% of respondents' scores were based on seven items or less. Only about 12% of respondents received the maximum number of items before the desired precision was reached.

### Short form development

The final IRT parameters in Table A1 were used to select items for the RAND-IAQL short forms. We determined that the impact of asthma on global aspects of QoL could be reliably assessed with as few as four carefully selected items, and four such items were selected to comprise the first short form (i.e., the RAND-IAQL-4; marginal reliability = 0.86). We next selected additional items to broaden the content coverage and increase precision of the RAND-IAQL-4. Within several key content domains identified by Eberhart et al. [3], the research team, using graphical illustrations of IRT-based item-level measurement precision, compared the relative utility of including a given item. From each domain the team selected the single item that tended to provide the most precision across the widest range of the latent construct. In total, we incorporated eight additional items (i.e., one item each reflecting content related to physical limitations, social concerns, inhaler awareness, health concerns, and sleep difficulties, and three general items with strong psychometric properties) to comprise the RAND-IAQL-12 (marginal reliability = 0.93). Appendix A1 contains



abbreviations of the RAND-IAQL-4 and RAND-IAQL-12 items; Appendix A2 provides score translation tables for both short forms.

### Score precision

In comparing the various RAND-IAQL scores to one another (i.e., RAND-IAQL-BANK, RAND-IAQL-CAT (i.e., computer adaptive test), RAND-IAQL-4, and RAND-IAQL-12), we found that RAND-IAQL-4 and -12 scores were highly intercorrelated ( $r = 0.96$ ) and highly correlated with RAND-IAQL-BANK scores ( $r = 0.93$  and  $0.97$ , respectively), indicating that both short forms adequately represent the underlying latent dimension.

The standard errors of the score estimates were computed to determine the precision of individual RAND-IAQL-4, RAND-IAQL-12, and RAND-IAQL-CAT scores. Fig. 4 provides a visual display of score precision and indicates that the RAND-IAQL-12 (RMSE = 2.63) outperforms the RAND-IAQL-CAT (RMSE = 3.09) at all but the most extreme score ranges (less one standard deviation below the mean). The precision of the RAND-IAQL-CAT is more comparable to the RAND-IAQL-4 (RMSE = 3.74), where the RAND-IAQL-4 actually provides more precision from about the mean (50) to one standard deviation above the mean (60; greater impact); however, for most other score ranges, and on average, the RAND-IAQL-CAT provides more precise score estimates.

### Preliminary validity of the RAND-IAQL

Finally, as a preliminary indication of validity, the RAND-IAQL is compared to the ACT. Table 2 presents mean score estimates of each type of RAND-IAQL score for each of three control categories derived from the ACT (poorly controlled, somewhat controlled, well controlled). Fig. 5 treats the ACT as a continuous measure and illustrates its relationship with the RAND-IAQL graphically. Both Table 2 and Fig. 5 indicate a similar pattern of findings. Not only do impact scores decrease with increasing control as hypothesized, but RAND-IAQL scores for those in the ACT “somewhat controlled” category are very close to the RAND-IAQL mean of 50, and the other two groups' means are symmetrically distributed around the mean of 50 (i.e.  $\sim \pm 7$ ).

### Discussion

In response to calls from leaders in the asthma research field, we developed an item bank and short forms to measure the perceived burden or impact of asthma on quality of life. The items reflect the impact of asthma in adults both on global QoL and on a wide range of specific QoL sub-domains and intentionally exclude items reflecting symptoms or activity levels.

One of the important questions we explored was whether asthma QoL is represented by several empirically distinct domains, or rather is more accurately characterized as a single, yet multi-faceted construct. Many existing measures of asthma-specific QoL have subscales (e.g., Jones et al., 1992; Juniper et al., 1999; Marks et al., 1992), but there is some controversy as to whether subscales are warranted [11,27–29]. Our findings suggest that, at least within this large and varied pool of items assessing the impact of asthma on QoL, specific subscales do not provide uniquely useful information. More specifically, despite the

inclusion of widely diverse content in our item bank, empirical evidence suggests that the set of items is best characterized as a unidimensional construct.

Our findings also demonstrate that the item bank, RAND-IAQL-4 and RAND-IAQL-12 each have excellent psychometric properties. For example, the full bank can precisely estimate the impact of asthma on QoL with reliability exceeding 0.90 in most places along the construct continuum; compared to the full item bank the RAND-IAQL-4 and RAND-IAQL-12 sacrifice very little measurement precision, despite drastically reducing respondent burden. However, both short forms provide less precision for scores reflecting less impact of asthma on QoL (e.g., scores one (40) or two (30) standard deviations below the mean). Fig. 5 illustrates this effect graphically by comparing scores for the RAND-IAQL item bank to ACT assessment of asthma control. Here we note a general floor effect where the RAND-IAQL is unable to provide score estimates for individuals more than two standard deviations below the mean (i.e., patients with little impact of asthma on QoL or well-controlled asthma). Similarly, a real-data computer adaptive test simulation demonstrated that adaptive assessment is comparable to or better than both short form assessments, performing slightly better than the RAND-IAQL-4, especially at the extreme ends of the continuum, with a very low average number of administered items (See Fig. 4). Finally, the high correlations among scores from the different assessment options support the assumption that they all measure the same underlying construct.

Importantly, our preliminary evaluation of the item bank's validity yielded encouraging results, with item bank scores varying with asthma control as expected. Additional validity analyses are underway that we expect will bolster this initial result by examining the item bank's scores in relation to other asthma outcome measures, demographic groupings and other health-related constructs. Though these preliminary validation efforts appear promising, the value of the item bank will be significantly increased through additional studies demonstrating its use in clinical populations, especially with respect to its sensitivity to change over time in response to treatment.

A well-designed item bank has several qualities that are attractive to users; most notable among them is that the bank allows users to create tailored assessments designed for a specific purpose. In the case of the RAND-IAQL item bank, for example, a researcher may be particularly interested in assessing individuals' standing in terms of social relationships, perhaps to inform an intervention design. This can be accomplished through careful selection of items with appropriate content, and because the items have been calibrated on the same underlying construct (i.e., the impact of asthma on QoL), scores from this tailored assessment can still be compared to scores from other item bank assessments (e.g., RAND-IAQL-4, RAND-IAQL-12, and RAND-IAQL-CAT).

Another attractive feature of the item bank is its versatility and sustainability. Now that the bank is in place, future research can be conducted to extend the bank in any number of interesting directions (e.g., to be relevant for children with asthma, for those with particular comorbid conditions, for Spanish speakers, etc.). Further, if the field's conceptual understanding of the impact of asthma on QoL shifts over time, items can be added and removed from the bank to accommodate those shifting paradigms. To enhance the

comparability of the bank to other widely used asthma outcome measures, interested users need only conduct a simple data collection using the bank (or a subset of the bank) and the additional outcome measure, and perform straightforward analyses to generate crosswalks between the two measures. For example, our research team is currently conducting analyses to provide one such crosswalk from the item bank to the Mark's Asthma Quality of Life Questionnaire [11].

This study has a number of important strengths including use of a large representative sample, rigorous methodology including modern psychometrics and qualitative item development informed by PROMIS® guidelines [30], and compelling evidence that impact of asthma on QoL can be reliably measured with just a few items. However, there are also some limitations. Most notably, the sample was recruited via an internet panel. Despite our success in achieving a demographically varied sample in terms of racial/ethnic distribution, it is possible that the respondents in our study are different from the general population of people with asthma in ways that we did not measure and cannot control for. This limitation underscores the importance of conducting follow-up studies with clinical samples and other more traditional research samples.

In response to calls from experts in the field, we employed state-of-the-art methods to develop a new assessment system for measuring the impact of asthma on quality of life. The RAND-IAQL item bank provides a psychometrically sound and versatile set of tools for measurement of the impact of asthma on quality of life in a way that is unconfounded with symptoms, control, and functioning. Assessments from the bank are highly reliable and minimally burdensome, and scores from different sets of bank items, chosen for specific purposes, can be directly compared across time and across studies. Interested users are encouraged to contact the first author to obtain the complete item bank, RAND-IAQL-12, and RAND-IAQL-4. Additionally, by the end of 2013 a computerized adaptive test version of the item bank will be available from [http://www.rand.org/health/surveys\\_tools/iaql.html](http://www.rand.org/health/surveys_tools/iaql.html).

## Acknowledgments

We would like to thank our expert panel which includes Eric Kleerup, MD (Clinical Professor, David Geffen School of Medicine, Division of Pulmonary and Critical Care Medicine), Steve Erickson, Ph.D. (Associate Professor at the University of Michigan College of Pharmacy), Cynthia Rand, Ph.D. (Professor of Medicine at the Johns Hopkins School of Medicine), Felita Jones (Asthma and Allergy Foundation of America), and Chris Draft (community activist and founder of the Chris Draft Family Foundation).

## Appendix

**Table A1**

RAND Impact of Asthma on QoL item bank IRT parameters and CAT exposure rates using actual cases.

Item stem <sup>a</sup>	Short form <sup>b</sup>	a	b1	b2	b3	b4	Exposure % <sup>c</sup>
Couldn't enjoy life	4/12	3.96	-0.20	0.48	1.01	1.52	99.2
Missed out on doing things with others	4/12	3.83	-0.27	0.38	0.90	1.40	45.9
Frustrated that have to do things differently than others	4/12	3.34	-0.37	0.32	0.88	1.39	0.0

Item stem <sup>a</sup>	Short form <sup>b</sup>	a	b1	b2	b3	b4	Exposure % <sup>c</sup>
Worry about asthma triggers	4/12	2.40	-0.90	0.14	0.84	1.49	20.9
Couldn't make plans in advance	12	3.81	0.16	0.62	1.07	1.62	13.3
Asthma controlling my life	12	3.64	-0.24	0.43	0.97	1.55	0.9
Everyday activities a struggle	12	3.48	-0.26	0.49	1.10	1.66	0.4
Asthma placed stress on relationships	12	3.27	0.18	0.72	1.26	1.73	2.8
Had to plan to make sure I always had an inhaler ready	12	2.66	-0.24	0.40	0.97	1.58	0.0
Asthma was on my mind	12	2.52	-0.89	0.23	0.91	1.58	24.4
Hard to get a good night's sleep	12	2.51	-0.38	0.43	1.02	1.59	0.0
Worried about long-term effects of asthma on my health	12	2.40	-0.79	0.05	0.71	1.32	15.6
Asthma interfered with my social life		3.90	-0.14	0.55	1.11	1.54	43.7
Felt bothered by limitations in what I could do		3.53	-0.50	0.26	0.80	1.37	49.8
Enjoyed the time I spent with others less		3.52	0.06	0.61	1.15	1.70	4.6
Felt bothered having to avoid situations or places		3.51	-0.27	0.42	0.99	1.57	0.0
Hard to do the things I enjoy doing.		3.50	-0.44	0.40	0.95	1.47	18.8
Asthma preventing me from achieving what I want in life		3.49	-0.13	0.47	0.99	1.49	0.1
Felt generally limited		3.46	-0.43	0.34	0.93	1.48	10.0
Cut back on things I enjoy		3.45	-0.27	0.44	0.97	1.56	0.0
Cannot do something without thinking about effect on asthma		3.45	-0.43	0.35	0.94	1.45	0.4
It bothered me that I have to plan ahead		3.33	-0.04	0.55	1.10	1.70	1.9
Kept from doing things I needed to do at work, school, or home		3.31	-0.24	0.50	1.07	1.69	1.5
Unable to do all the things I wanted to do		3.25	-0.50	0.34	0.94	1.45	13.2
Asthma affected my life more than I want to admit		3.24	-0.44	0.30	0.89	1.44	0.0
Managing asthma took effort		3.21	-0.32	0.42	0.99	1.56	0.0
Felt I could not control my asthma		3.11	-0.14	0.54	1.11	1.66	0.0
Bothered at work, school, or home		3.10	-0.48	0.42	1.05	1.64	0.0
Felt different than other people		3.08	-0.15	0.49	1.02	1.56	0.0
Had to be careful what I did		3.08	-0.74	0.23	0.85	1.46	32.5
Because of asthma I felt helpless		3.03	-0.22	0.33	1.10	1.72	0.0
Worried I would have an attack while visiting a new place		2.90	-0.24	0.43	1.01	1.58	0.0
Did things for shorter amounts of time than I would have liked		2.86	-0.62	0.22	0.83	1.51	24.7
Found myself making excuses to others		2.83	0.16	0.69	1.23	1.80	1.4
Other people didn't understand my asthma		2.83	0.06	0.56	1.10	1.66	0.0
Avoided situations where my asthma might embarrass me		2.79	0.05	0.56	1.08	1.67	0.0
Asthma interfered with romantic relationships		2.77	0.22	0.73	1.25	1.76	0.0
Worried about asthma attack in front of others		2.74	-0.10	0.48	1.00	1.47	0.0
Frustrated that I can't control the things that trigger asthma		2.73	-0.54	0.28	0.85	1.40	0.0

Item stem <sup>a</sup>	Short form <sup>b</sup>	a	b1	b2	b3	b4	Exposure % <sup>c</sup>
Bothered that have to be aware of possible asthma triggers		2.72	-0.65	0.18	0.82	1.43	17.0
Bothered that don't know when my asthma will get worse.		2.71	-0.64	0.16	0.79	1.35	13.4
Avoiding triggers created problems in my relationships		2.70	0.07	0.66	1.28	1.83	1.3
Because of asthma I felt anxious.		2.69	-0.56	0.11	1.06	1.80	0.9
Can't visit friends or family because of triggers in their home		2.67	0.01	0.54	1.14	1.71	0.0
Worried about dying from an attack		2.67	-0.02	0.51	0.98	1.41	0.0
Because of my asthma I felt irritable		2.63	-0.57	0.05	1.04	1.81	5.1
Felt frustrated that I can't fix or get away from my asthma		2.59	-0.67	0.17	0.70	1.23	12.5
Hard having to speak up about things that trigger my asthma		2.53	0.00	0.55	1.07	1.72	0.0
Afraid to be physically active		2.52	-0.62	0.21	0.82	1.46	11.7
Felt scared when an attack came on		2.48	-0.41	0.27	0.85	1.37	0.0
Worried about taking daily asthma medications		2.43	0.00	0.59	1.16	1.81	0.0
It was annoying having to have enough medication on hand		2.42	-0.17	0.48	1.10	1.65	0.0
Asthma bothered people I care about		2.41	0.06	0.69	1.25	1.82	0.6
Worried about becoming immune to my medication		2.26	-0.22	0.43	1.07	1.64	0.0
I struggled with the pros and cons of taking asthma medication		2.23	-0.07	0.53	1.21	1.85	0.0
Asthma kept me from having things I wanted		2.20	-0.22	0.39	0.98	1.57	0.0
Had to make compromises because of treatment costs		2.15	-0.05	0.53	1.16	1.74	0.0
Worried about using too much medication		2.15	-0.04	0.53	1.16	1.78	0.0
Bothered by the way medication made me feel		2.14	-0.01	0.66	1.29	1.87	0.0
Worried that medications will make future health worse		2.13	-0.32	0.38	1.09	1.73	0.0
People thought my asthma symptoms were cold symptoms		2.10	0.16	0.68	1.29	1.96	0.7
Felt dependent on my medication		2.03	-0.57	0.27	0.86	1.47	0.0
Worried about not having my inhaler when I need it		2.00	-0.62	0.25	0.85	1.44	6.6
Worried about getting a cold with my asthma		1.83	-0.55	0.19	0.85	1.51	0.0
Annoying having to carry my inhaler with me		1.81	-0.01	0.66	1.36	1.94	0.0

Combined sample ( $N = 2032$ ) fit indices:  $\chi^2 = 16,559$ ,  $df = 1710$ ; CFI = 0.965; TLI = 0.964; RMSEA = 0.065.

Notes.

<sup>a</sup> Actual item wording has been abbreviated.

<sup>b</sup> This column indicates items that appear in both the RAND-IAQL-4 and RAND-IAQL-12 ("4/12") and those items that appear in only the RAND-IAQL-12 ("12").

<sup>c</sup> Exposure % indicates the percentage of the  $N = 2032$  real-data CATs that were administered a given item.

**Table A2**

RAND-IAQL-4 and RAND-IAQL-12 sum score to IRT-score translations.

Sum score	EAP	SE	Observed frequency
RAND-IAQL-12			
0	32.7	5.1	8.5
1	37.4	3.7	6.3
2	39.9	3.1	5.1
3	41.6	2.8	4.4
4	43.0	2.6	5.2
5	44.2	2.4	4.0
6	45.3	2.2	4.1
7	46.2	2.1	3.1
8	47.1	2.0	3.2
9	47.9	2.0	2.7
10	48.6	1.9	2.4
11	49.3	1.9	3.0
12	49.9	1.9	2.6
13	50.6	1.8	2.1
14	51.2	1.8	2.2
15	51.8	1.8	1.6
16	52.4	1.8	1.8
17	52.9	1.8	2.3
18	53.5	1.8	1.6
19	54.0	1.8	1.4
20	54.5	1.7	1.3
21	55.1	1.7	1.4
22	55.6	1.7	1.6
23	56.1	1.7	1.7
24	56.6	1.7	1.7
25	57.1	1.7	1.7
26	57.6	1.7	1.9
27	58.1	1.7	1.5
28	58.6	1.7	1.4
29	59.1	1.7	1.6
30	59.6	1.7	1.6
31	60.1	1.7	1.4
32	60.6	1.7	1.3
33	61.1	1.7	0.9
34	61.7	1.7	1.1
35	62.2	1.7	0.9
36	62.7	1.7	0.9
37	63.3	1.8	1.0

Sum score	EAP	SE	Observed frequency
RAND-IAQL-12			
38	63.9	1.8	0.7
39	64.5	1.8	1.0
40	65.1	1.9	0.6
41	65.8	1.9	0.6
42	66.6	2.0	0.8
43	67.4	2.1	0.6
44	68.3	2.2	0.8
45	69.3	2.4	0.5
46	70.6	2.6	0.5
47	72.2	2.9	0.3
48	75.6	4.1	1.0
RAND-IAQL-4			
0	36.2	5.7	16.9
1	41.8	4.0	13.2
2	45.1	3.4	9.2
3	47.5	3.1	8.2
4	49.4	3.0	7.2
5	51.1	2.9	4.7
6	52.7	2.8	5.6
7	54.2	2.8	4.7
8	55.6	2.8	5.0
9	57.0	2.7	4.6
10	58.4	2.7	3.7
11	59.8	2.7	4.1
12	61.3	2.8	3.2
13	62.9	2.9	2.6
14	64.7	3.1	2.5
15	66.9	3.3	1.8
16	71.3	4.6	2.9

Note: The “EAP” column is in the T-score metric (mean = 50, standard deviation = 10). Sum score refers to the sum of the item scores (0–4), so for the RAND-IAQL-4 the range of possible scores is 0–16; for the RAND-IAQL-12 the range of possible scores is 0–48. When scoring the RAND-IAQL-4, we recommend only scoring individuals with complete, non-missing responses. For the RAND-IAQL-12, users may impute the mean item response when the number of missing responses is five or less.

## References

1. National Heart Lung, and Blood Institute. Expert panel report 3: guidelines for the diagnosis and management of asthma. (US): Bethesda, MD: National Heart, Lung, and Blood Institute; 2007. National asthma education and prevention program, third expert panel on the diagnosis and management of asthma.
2. Wilson SR, et al. Asthma outcomes: quality of life. *J Allergy Clin Immunol.* 2012; 129(3):S88–123. [PubMed: 22386511]

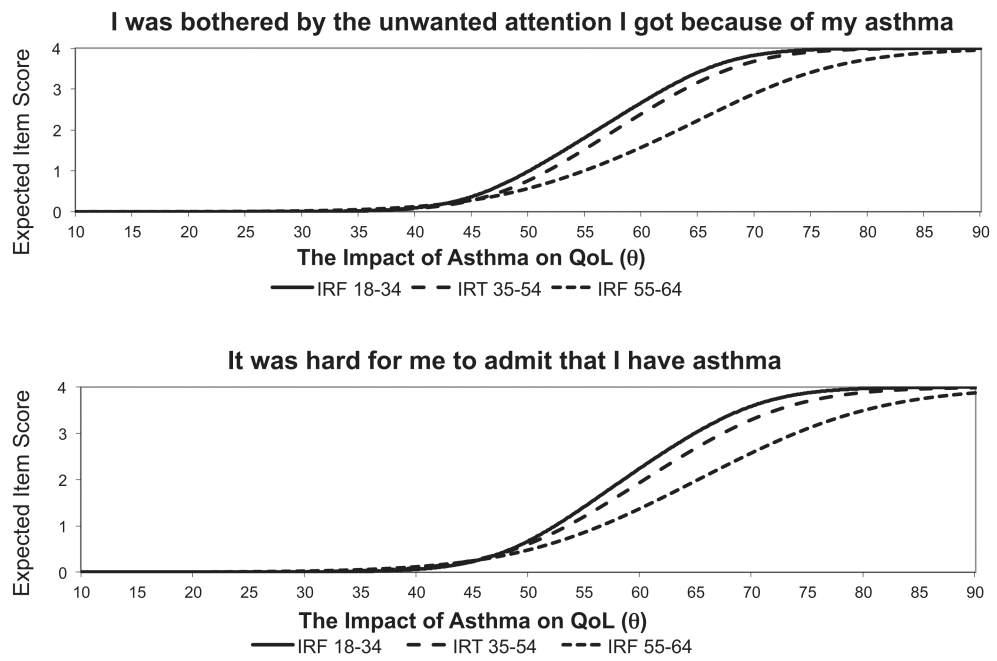
3. Eberhart, NK.; Sherbourne, CD.; Edelen, MO.; Stucky, B.; Lara, M. Development of a measure of asthma-specific quality of life among adults. *Qual Life Res.* 2013. <http://dx.doi.org/10.1007/s11136-013-0510-x>
4. Reeve BB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007; 45(5 Suppl. 1):S22–31. [PubMed: 17443115]
5. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monogr Suppl.* 1969; 34(4, Pt. 2):100.
6. CDC. America breathing easier 2010: CDC's national asthma control program at a glance. National Center for Environmental Health (NCEH): Division of Environmental Hazards and Health Effects. 2010:4.
7. Adams R, et al. Assessment of an asthma quality of life scale using item-response theory. *Respirology.* 2005; 10(5):587–93. [PubMed: 16268911]
8. Barley EA, Quirk FH, Jones PW. Asthma health status measurement in clinical practice: validity of a new short and simple instrument. *Respir Med.* 1998; 92(10):1207–14. [PubMed: 9926151]
9. Benninger MS, Senior BA. The development of the Rhinosinusitis Disability Index. *Arch Otolaryngol Head Neck Surg.* 1997; 123(11):1175–9. [PubMed: 9366696]
10. Hyland ME, et al. Measurement of psychological distress in asthma and asthma management programmes. *Br J Clin Psychol.* 1995; 34(Pt 4):601–11. [PubMed: 8563667]
11. Marks GB, Dunn SM, Woolcock AJ. A scale for the measurement of quality of life in adults with asthma. *J Clin Epidemiol.* 1992; 45(5):461–72. [PubMed: 1588352]
12. Sibbald B, Collier J, D'Souza M. Questionnaire assessment of patients' attitudes and beliefs about asthma. *Fam Pract.* 1986; 3(1):37–41. [PubMed: 3956902]
13. Tu SP, et al. A new self-administered questionnaire to monitor health-related quality of life in patients with COPD. Ambulatory Care Quality Improvement Project (ACQUIP) investigators. *Chest.* 1997; 112(3):614–22. [PubMed: 9315792]
14. Yeatts KB, et al. Construction of the pediatric asthma impact scale (PAIS) for the patient-reported outcomes measurement information system (PROMIS). *J Asthma.* 2010; 47(3):295–302. [PubMed: 20394514]
15. Hyland ME. The living with asthma questionnaire. *Respir Med.* 1991; 85(Suppl. B):13–6. discussion 33-7. [PubMed: 1759015]
16. Steinberg L. Context and serial-order effects in personality measurement - limits on the generality of measuring changes the measure. *J Pers Soc Psychol.* 1994; 66(2):341–9.
17. Nathan RA, et al. Development of the asthma control test: a survey for assessing asthma control. *J Allergy Clin Immunol.* 2004; 113(1):59–65. [PubMed: 14713908]
18. Schatz M, et al. Validation of the asthma impact survey, a brief asthma-specific quality of life tool. *Qual Life Res.* 2007; 16(2):345–55. [PubMed: 17033905]
19. Muthén, LK.; Muthén, BO. Mplus user's guide: statistical analysis with latent variables. 6th. Los Angeles: Muthén & Muthén; 2010. p. 1online resource (752 S
20. Hu LT, Bentler P. 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.
21. Cai, L.; du Toit, SHC.; Thissen, D. IRTPRO Version 2: flexible, multidimensional, multiple categorical IRT modeling. Chicago, IL: Scientific Software International; 2011. Computer software
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate -a practical and powerful approach to multiple testing. *J R Stat Soc Series B Methodol.* 1995; 57(1):289–300.
23. Raju NS. The area between 2 item characteristic curves. *Psychometrika.* 1988; 53(4):495–502.
24. Irwin DE, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Qual Life Res.* 2010; 19(4):595–607. [PubMed: 20213516]
25. DeWitt EM, et al. Construction of the eight-item patient-reported outcomes measurement information system pediatric physical function scales: built using item response theory. *J Clin Epidemiol.* 2011; 64(7):794–804. [PubMed: 21292444]



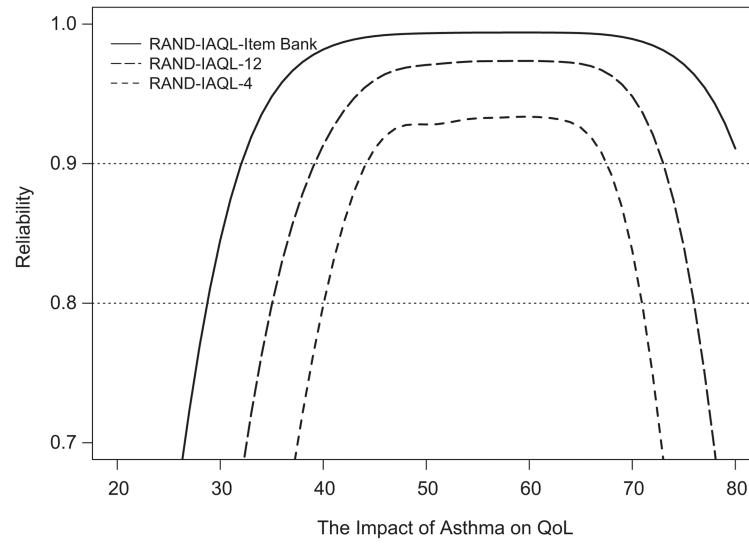
26. Thissen, D.; Wainer, H. Test scoring. Mahwah, N.J: L. Erlbaum Associates; 2001. p. 1online resource (xii, 422 p.)
27. Puhon MA, et al. Comparing a disease-specific and a generic health-related quality of life instrument in subjects with asthma from the general population. *Health Qual Life Outcomes*. 2008; 6
28. Juniper EF, et al. Validation of a standardized version of the asthma quality of life questionnaire. *Chest*. 1999; 115(5):1265–70. [PubMed: 10334138]
29. Jones PW, et al. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *Am Rev Respir Dis*. 1992; 145(6):1321–7. [PubMed: 1595997]
30. DeWalt DA, et al. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care*. 2007; 45(5 Suppl. 1):S12–21. [PubMed: 17443114]

## Abbreviations

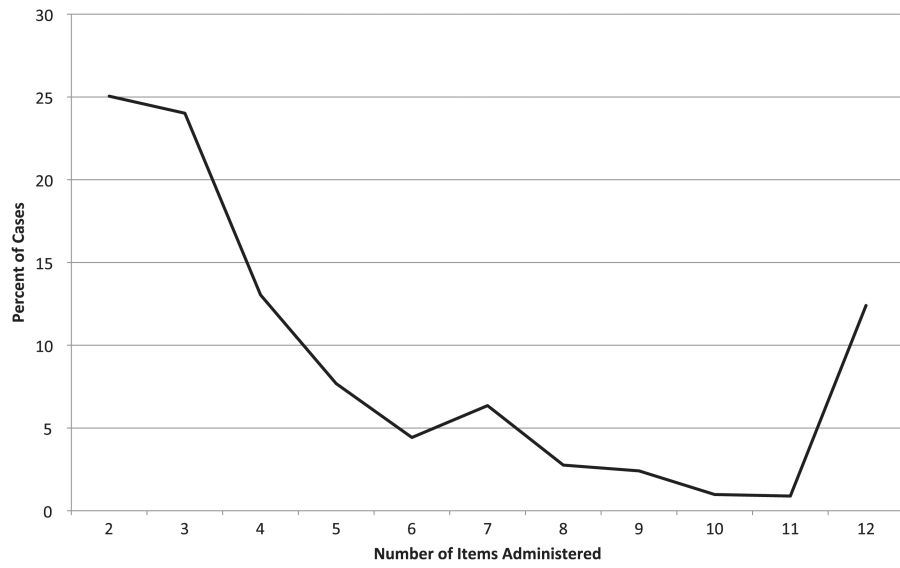
<b>ACT</b>	asthma control test
<b>COPD</b>	chronic obstructive pulmonary disease
<b>CFA</b>	confirmatory factor analysis
<b>DIF</b>	differential item functioning
<b>EFA</b>	exploratory factor analysis
<b>GRM</b>	graded response model
<b>IRT</b>	item response theory
<b>RAND-IAQL</b>	RAND impact of asthma on quality of life
<b>RMSE</b>	root mean square error
<b>NHLBI</b>	National Heart, Lung, and Blood Institute
<b>PROMIS</b>	Patient reported outcomes measurement information system
<b>QoL</b>	Quality of life
<b>wABC</b>	Weighted area between the expected score curves
<b>WLSMV</b>	Mean and variance adjusted weighted least squares



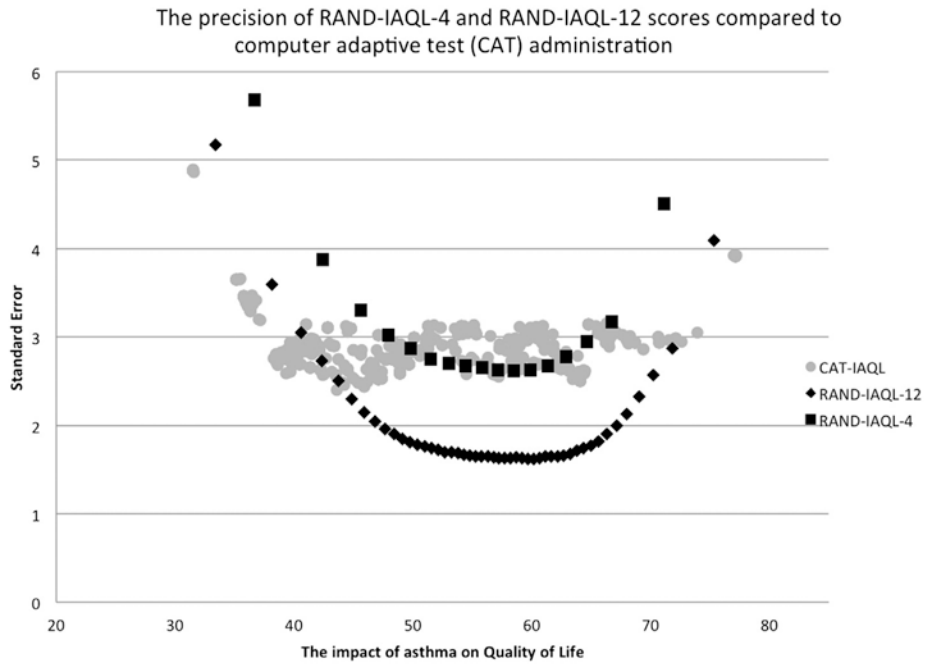
**Figure 1.**  
Two examples of items removed for differential item functioning.



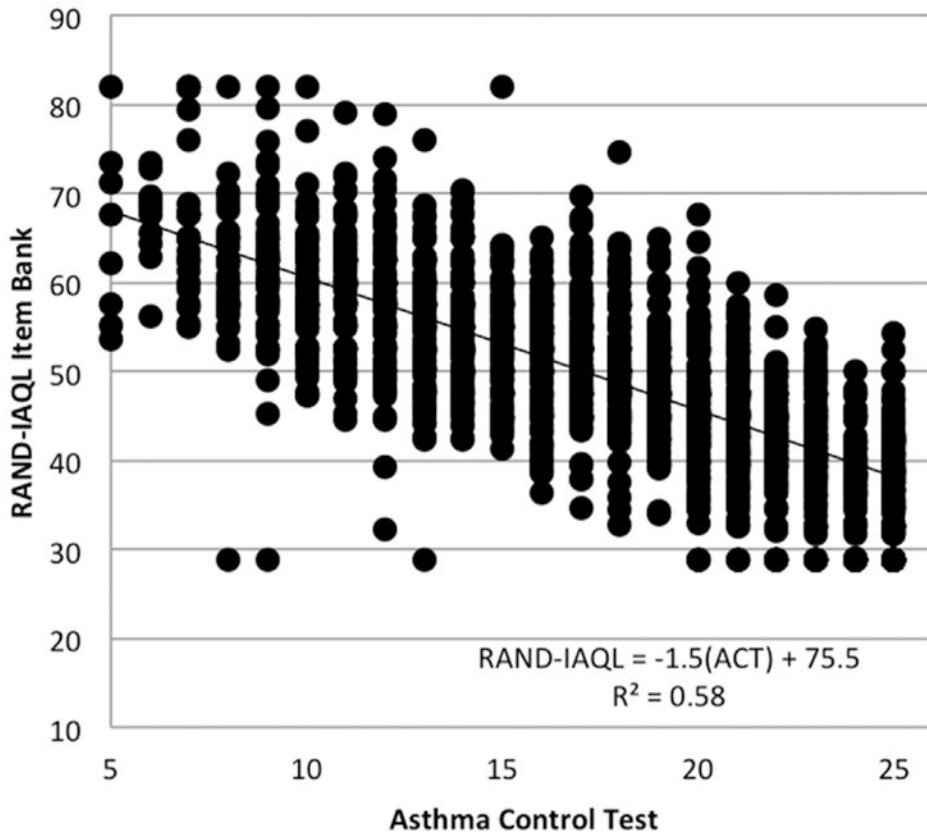
**Figure 2.**  
Reliability of the full bank, RAND-IAQL-4, and RAND-IAQL-12.



**Figure 3.**  
Percent of cases ( $N = 2032$ ) receiving each possible item count.



**Figure 4.** The precision of RANG-IAQL-4 and RANG-IAQL-12 scores compared to computer adaptive test (CAT) administration.



**Figure 5.**

Better asthma control is related to less impact of asthma on QoL. The x-axis indicates scores on the asthma control test (ACT) that range from 5 to 25 (i.e., the sum of five items each scored 1–5). Higher scores on the ACT indicate improvements in asthma control (scores greater than 19 are commonly associated well-controlled asthma). The y-axis indicates scores on the RAND-Impact of asthma on quality of life item bank (RAND-IAQL). Scores are arranged on a standardized T-score metric with a mean of 50 and a standard deviation of 10. For example RAND-IAQL scores of 60 and 70 indicate scores one and two standard deviations above the mean, respectively. High scores on the RAND-IAQL item bank indicate negative impact of asthma on quality of life (i.e., worsening QoL).

**Table 1**

Characteristics of the exploratory and confirmatory samples.

	Exploratory group (N = 1500)	Confirmatory group (N = 532)
Female %	60	61
Age mean (SD; range)	43.3 (14.9; 18–86)	43.0 (14.3; 18–77)
Race/Ethnicity, %		
Non-Hispanic White	81	80
African American	19	23
Hispanic	10	14
Asian	12	9
Other	3	1
Education %		
<High school graduate	3	3
High school graduate	15	14
<BA/BS degree	36	38
BA/BS degree	25	22
Graduate degree	22	23
Employment %		
Full-time	50	53
Part-time	10	8
Not employed	17	18
Retired/student/homemaker	23	21
Income %		
<\$25,000	19	19
\$25,000–\$49,999	26	22
\$50,000–\$99,999	31	34
>\$100,000	24	25

**Table 2**

Comparisons of RAND impact of asthma on quality of life (RAND-IAQL) scores by the asthma control test (ACT).

RAND impact of asthma on QoL (RAND-IAQL)	Asthma control test level		
	Poorly controlled ( <i>N</i> = 796)	Somewhat controlled ( <i>N</i> = 446)	Well controlled ( <i>N</i> = 770)
RAND-IAQL-Full Item Bank	57.4 (7.9)	49.9 (6.5)	41.1 (7.3)
RAND-IAQL-CAT	56.8 (7.6)	49.7 (7.1)	41.8 (7.2)
RAND-IAQL-12	57.1 (7.5)	49.7 (6.6)	41.6 (6.4)
RAND-IAQL-4	56.4 (8.0)	49.4 (7.2)	42.7 (6.2)

Note: Sample sizes were slightly smaller for the RAND-IAQL-4 and RAND-IAQL-12 due to missing data (*N* = (775, 751); *N* = (437, 429); and *N* = (766, 741)) for the categories Poorly Controlled, Somewhat Controlled, and Well Controlled, respectively.