

RESEARCH

Open Access



# Psychometric properties of computerized adaptive testing for chronic obstructive pulmonary disease patient-reported outcome measurement

Jiajia Wang<sup>1,2,3†</sup>, Yang Xie<sup>1,2,3†</sup>, Zhenzhen Feng<sup>1,2,3</sup> and Jiansheng Li<sup>1,2,3\*</sup>

## Abstract

**Background** Computerized adaptive testing (CAT) is an effective way to reduce time, repetitious redundancy, and respond burden, and has been used to measure outcomes in many diseases. This study aimed to develop and validate a comprehensive disease-specific CAT for chronic obstructive pulmonary disease (COPD) patient-reported outcome measurement.

**Methods** The discrimination and difficulty of the items from the modified patient-reported outcome scale for COPD (mCOPD-PRO) were analyzed using item response theory. Then the initial item, item selection method, ability estimation method, and stopping criteria were further set based on *Concerto* platform to form the CAT. Finally, the reliability and validity were validated.

**Results** The item discrimination ranged from 1.05 to 2.71, and the item difficulty ranged from  $-3.08$  to 3.65. The measurement reliability of the CAT ranged from 0.910 to 0.922 using random method, while that ranged from 0.910 to 0.924 using maximum Fisher information (MFI) method. The content validity was good. The correlation coefficient between theta of the CAT and COPD assessment test and modified Medical Research Council dyspnea scale scores using random method was 0.628 and 0.540 ( $P < 0.001$ ;  $P < 0.001$ ) respectively, while that using MFI method was 0.347 and 0.328 ( $P = 0.007$ ;  $P = 0.010$ ) respectively. About 11 items (reducing by 59.3%) on average were tested using random method, while about seven items (reducing by 74.1%) on average using MFI method. The correlation coefficient between theta of the CAT and mCOPD-PRO total scores using random method was 0.919 ( $P < 0.001$ ), while that using MFI method was 0.760 ( $P < 0.001$ ).

**Conclusions** The comprehensive disease-specific CAT for COPD patient-reported outcome measurement is well developed with good psychometric properties, which can provide an efficient, accurate, and user-friendly measurement for patient-reported outcome of COPD.

<sup>†</sup>Jiajia Wang and Yang Xie have contributed equally to this work.

\*Correspondence:  
Jiansheng Li  
li\_js8@163.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

**Keywords** Chronic obstructive pulmonary disease, Patient-reported outcome, Computerized adaptive testing, Item response theory, Reliability, Validity

## Background

Chronic obstructive pulmonary disease (COPD), characterized by chronic respiratory symptoms and persistent (often progressive) airflow limitation, is a leading cause of morbidity and mortality worldwide inducing an economic and social burden that is both substantial and increasing [1]. The prevalence of COPD among people aged 40 years or older is 12.64% around the whole world, and 13.7% in China [2, 3]. COPD is the third leading causes of death responsible for approximately 6% of the world's total deaths in 2019 [4]. Health status in patients with COPD declines over time [5]. A patient-reported outcome (PRO) is any report of a patient's health status derived directly from the patient [6]. It has been increasingly recognized that PRO instruments play an important role in assessing health status and treatment outcome of COPD [7, 8]. Previously, our team has developed and validated a 27-item comprehensive disease-specific PRO measurement, the modified Patient-reported Outcome Scale for COPD (mCOPD-PRO), which is a 5-point Likert scale including physiological, psychological, and environmental domains, with lower scores indicating better health status [9]. In the stage of validation of the instrument, the measurement properties assessed referred to internal consistency reliability, content validity, construct validity, criterion validity, known groups validity, and feasibility [9]. It is showed that the mCOPD-PRO has good psychometric properties with the Cronbach's alpha of 0.954 [9]. As a common chronic disease worldwide, patients with COPD usually require frequent outcome measures. In this case, the response burden of measurement tools can't be ignored. Although the median completion time of the mCOPD-PRO is only 5 min, given the number of items, the response burden still needs to be considered [9].

Computerized adaptive testing (CAT), based on item response theory (IRT), is a form of testing that uses a computer to automatically select appropriate items for the examinee [10]. Generally speaking, CAT selects an item from an item bank that is appropriate to the examinee's theta, an index of the latent trait in IRT, here the health status, and then updates the examinee's theta according to the responses to this item [10]. This process is repeated until the examinee's theta is accurately estimated [10]. CAT is an effective way to reduce time, repetitious redundancy, and respond burden, and has been increasingly used for psychological and health measurement in an efficient, accurate, and user-friendly manner [11–13]. Recently, CAT has also been used for outcome measurement of COPD [14–22]. Norweg A

et al. developed a modified and expanded item bank of Dyspnea Management Questionnaire measuring the dyspnea, one of the most prominent symptoms in COPD, to construct the CAT using an internally developed CAT program, and subsequently conducted CAT real data simulations to estimate the CAT's accuracy, precision, and validity [14]. Choi SW et al. and Yount SE, et al. also focused on the dyspnea in COPD, and developed a measure of dyspnea and related functional limitations, where post-hoc CAT simulations were conducted to determine the number of items required to achieve high precision [15, 16]. Additionally, Yount SE et al. administered the Patient-Reported Outcomes Measurement Information System (PROMIS®) measures using CAT, followed by administration of any remaining short form items that had not yet been administered by CAT, to examine their responsiveness to changes associated with COPD exacerbation recovery [17]. Paap MCS et al. performed simulations using empirical data from patients with COPD to assess the incremental value of multidimensional CAT compared with unidimensional CAT, and investigated the item usage for the multidimensional CAT drawing items from three PROMIS domains (fatigue, physical function, and ability to participate in social roles and activities) and a COPD-specific item bank [18, 19]. O'Hoski S et al. estimated the test-retest reliability, construct validity, and responsiveness of the CAT version of Late Life Disability Instrument in patients with COPD, which was a participation measure not specifically developed for COPD [20, 21]. Ho EH et al. used plausible values to account for measurement error and analyze the probability of true within-individual change in a sample of patients with COPD completing two PROMIS domains (physical function and fatigue), and indicated that CAT have better ability to detect change compared to short forms [22]. However, although two of the measures are designed for COPD, they are not comprehensive. By comparison, the others are multi-dimensional, but they are not designed for COPD or only include part of disease-specific items. The comprehensive disease-specific CAT for PRO measurement of COPD is urgently needed. Therefore, the current study aimed to develop and validate a comprehensive disease-specific CAT for PRO measurement of COPD based on the paper-and-pencil version of mCOPD-PRO.

## Methods

### Developing the CAT

The CAT was developed based on an open-source online adaptive testing platform *Concerto* developed and

maintained by the University of Cambridge Psychometrics Centre (Available from: <https://concertoplatform.com/about>) [23]. The essential elements of the CAT involved item parameters calibration, initial item, item selection method, ability estimation method, and stopping criteria [24].

#### Item parameters calibration

The item bank of the CAT was the same as the paper-and-pencil version of mCOPD-PRO. Item parameters including discrimination (a) and difficulty (b) were estimated using graded response model of IRT. Before modelling, the core assumptions (unidimensionality, local independence, and monotonicity) were evaluated. The unidimensionality assumption was assessed using both exploratory and confirmatory factor analysis. For the former, a ratio of the eigenvalue for the first factor to the second factor in excess of four was supportive of the unidimensionality assumption [25]. As for the latter, the recommended criteria for fit indices were as follows: (1) the comparative fit index and non-normed fit index were close to 0.90; (2) the incremental fit index was close to 0.95; and (3) the standardized root-mean-square residual and root-mean-square error of approximate (RMSEA) were close to 0.08 [26, 27]. A Chen and Thissen's index of  $>0.30$  implied possible local dependence [28]. The monotonicity was considered acceptable if the scalability coefficients ( $H_i$ ) for the items were  $>0.30$  [29]. The data were from 366 patients with COPD in the phase of validation of the paper-and-pencil version of mCOPD-PRO [9]. The mean age was 66 years; 279 cases (76.2%) were males and 87 cases (23.8%) were females.

#### Initial item and item selection method

Item selection is dependent on the examinee's responses to a given item. The random and maximum Fisher information (MFI) methods were adopted to select items. The former is simple and effective, and selects items randomly, while the latter selects items based on the information of items. Thus, the initial item selected by the random method is random, while that selected by the MFI method is the one with the maximum information.

#### Ability estimation method

As a key component of CAT, the ability estimation method not only affects the accuracy of ability estimation, but also affects the efficiency of item selection and the determination of stopping rule. In this study, the maximum likelihood estimation, one of the most widely used ability estimation method, was used to conduct ability estimation.

#### Stopping criteria

Generally speaking, the stopping criteria of CAT involves fixed length and standard error of measurement (SEM). The measurement accuracy may vary among subjects with different fixed lengths, and the fixed SEM can best reflect the core idea of CAT. In this study, the SEM of  $\leq 0.30$  was determined as our stopping criteria, which meant that the test was terminated if the pre-specified value of SEM was met or the item bank was exhausted [10].

#### Simulation test

The CAT simulation was performed to determine the appropriate sample size for the validation of the CAT based on the true participants' responses data from the phase of validation of the paper-and-pencil version of mCOPD-PRO [9]. R commands for the simulated data of different sample sizes (60 cases, 100 cases, 300 cases, 500 cases, 1000 cases, 3000 cases, and 5000 cases, respectively) were generated using Firestar version 1.5.1, and then ran in R version 3.5.1. The CAT simulation settings for Firestar version 1.5.1 were as follows: (1) for the IRT model, graded response model was used; (2) for the item selection method, random and MFI methods were used respectively; (3) for the stopping criteria, the maximum SEM was 0.30; and (4) for the simulated data, the mean was specified as zero with the standard deviation of one.

#### Validating the CAT

There is no consensus regarding the validation of CAT for PRO instruments. In our study, the reliability was estimated based on IRT method, while the content and criterion validity were evaluated referring to classical test theory method. It was assumed that the response to each item of the CAT was consistent with that of the paper-and-pencil version of mCOPD-PRO in the phase of validation [9]. However, the number of items selected by the CAT might be less than that of the paper-and-pencil version of mCOPD-PRO.

#### Reliability

The measurement reliability ( $r$ ) was calculated through the SEM of the CAT. The relationship between SEM and measurement reliability ( $r$ ) is inversely proportional function (assuming that the mean ability of the subjects is zero and the standard deviation is one). The formula:  $SEM = (1-r)^{1/2}$ ; that is, the measurement reliability ( $r = 1 - SEM^2$ ) [10, 24].

#### Content validity

The content validity of the CAT was assessed based on the content validity of the paper-and-pencil version of mCOPD-PRO in the phase of validation [9].

### Criterion validity

The COPD assessment test and modified Medical Research Council dyspnea scale (mMRC) recommended by the *Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease* were both used as gold standard [1, 30, 31]. The criterion validity was evaluated using correlation coefficient between the test result theta of the CAT and COPD assessment test and mMRC scores. The correlation coefficient of  $\geq 0.40$  was considered acceptable [9].

### Statistical analysis

Continuous data were expressed as the mean  $\pm$  standard deviation or median (interquartile range), while categorical data were presented as frequencies (percentages). The confirmatory factor analysis was conducted using LISREL version 8.70 (Scientific Software International, Inc., Chicago, IL, USA). The local independence and monotonicity were tested using “TestAnaAPP” and “mokken” packages in R (The R Foundation, Vienna, Austria), respectively. The IRT analysis was performed using MULTILOG version 7.03 (Scientific Software International, Inc., Chicago, IL, USA), and the CAT simulation was conducted using Firestar version 1.5.1 (Northwestern University Feinberg School of Medicine, Chicago, IL, USA) and R version 3.5.1 (The R Foundation, Vienna, Austria). Additionally, descriptive statistics, exploratory factor analysis, calculation of measurement reliability ( $r$ ), correlation analysis, and independent sample  $t$ -test were conducted using SPSS version 22.0 (IBM Corporation, Armonk, NY, USA).

## Results

### Calibration of item parameters

The exploratory factor analysis showed that the ratio of the eigenvalues for the first factor (12.551) to the second factor (2.015) is 6.229 surpassing the thresholds of 4. Moreover, the confirmatory factor analysis showed that comparative fit index, incremental fit index, non-normed fit index, standardized root-means-square residual, and RMSEA were 0.91, 0.91, 0.90, 0.11, and 0.16, respectively. The fit indices were close to the thresholds, except for the RMSEA which was less satisfactory. 32 of 351 item pairs (9.1%) showed a Chen and Thissen's index of above threshold of 0.30 with a maximum of 0.526, and 25 of 32 item pairs (78.1%) ranged from 0.30 to 0.40. Most of the locally dependent items were found to be related to respiratory symptoms of COPD, such as cough, sputum, chest tightness, panting, and shortness of breath. Given the fact that multiple respiratory symptoms often coexist, and that one symptom may involve more than one item (for example, the items “Did you cough?”, “Was your cough aggravated by daily activities?”, and “Was your cough aggravated by mood swings?”), the test results of

local independence are considered acceptable. The scalability coefficients ( $H_i$ ) for all the items ranged from 0.36 to 0.54 exceeding the thresholds of 0.30. Overall, it can be considered that all three assumptions for IRT analysis are met. The item discrimination (a) of mCOPD-PRO ranged from 1.05 to 2.71, and the item difficulty (b) ranged from  $-3.08$  to 3.65 (Table 1). Moreover, the difficulty (b) of 24 items (88.9%) was between  $-3.0$  and 3.0. The first and fifth item characteristic curve showed monotonous changes, and the second, third and fourth item characteristic curve showed normal distribution except for individual items (Additional file 1). The maximum value of total information was 34.224 (Additional file 2).

### Establishment of CAT model

The CAT model established based on *Concerto* platform included item bank, algorithm, test system, score report, and management system modules.

### Formation of the CAT

The CAT was formed. At least five links (*Login*, *Run*, *Test*, *Feedback*, and *End*) were needed to complete the whole test. Taking a test with random item selection method as an example, a total of 11 items were tested. The test procedure was presented in Additional file 3.

### Simulation test

The results of CAT simulation using random and MFI methods were described in Table 2, Figs. 1 and 2, and Additional file 4–9. Despite different sample sizes of simulated data (60 cases, 100 cases, 300 cases, 500 cases, 1000 cases, 3000 cases, and 5000 cases, respectively), the items administered in different simulation testes using random method were relatively discrete, and the average number was all ten. The correlation coefficient between simulated and true theta estimates ranged from 0.970 to 0.976, and the average SEM was 0.290 or 0.291. By comparison, the items administered in different simulation testes using MFI method were relatively centralized, and the average number was all seven. The correlation coefficient between simulated and true theta estimates ranged from 0.968 to 0.979, and the average SEM ranged from 0.289 to 0.292.

### Validation of the CAT

According to the results of CAT simulation above, the sample size for the validation of the CAT was determined as 60 cases. Therefore, the true participants' responses data from 60 patients with COPD in the phase of validation of the paper-and-pencil version of mCOPD-PRO were used to evaluate the reliability and validity of the CAT. The mean age was 65 years; 41 cases (68.3%) were males and 19 cases (31.7%) were females.

**Table 1** The item discrimination (a) and difficulty (b) of the modified patient-reported outcome scale for chronic obstructive pulmonary disease (mCOPD-PRO)

Item No.	Item description	a	b			
			b1	b2	b3	b4
1	Did you cough?	1.52	-2.80	-0.36	0.72	2.54
2	Did you have phlegm?	1.30	-2.84	-0.81	0.61	2.75
3	Did you have chest tightness?	1.92	-1.87	-0.75	0.26	2.00
4	Did you have panting?	2.18	-2.20	-0.92	0.05	1.67
5	Did you get short of breath?	2.16	-2.35	-0.86	0.06	1.85
6	Did you feel weak when you talked?	2.18	-1.22	-0.11	0.78	2.07
7	Did you have fatigue?	2.28	-1.74	-0.53	0.27	2.22
8	Did you feel that you were in a poor spirit?	2.07	-1.12	-0.33	1.01	2.33
9	Did you notice cyanosis (your lips, nails, skin, etc. turned purple)?	1.05	-0.54	0.80	2.23	3.43
10	Were you afraid of the wind (cold)?	1.22	-2.04	-0.93	0.37	2.36
11	Did you have abdominal distension?	1.18	-1.00	0.46	1.68	3.65
12	Did you often catch cold?	1.33	-2.74	-0.44	0.71	2.70
13	Were you able to go outdoors alone? (walking, going shopping, having a haircut, taking a bus, etc.)	1.90	-1.21	0.04	1.15	2.37
14.1	Was your cough aggravated by daily activities?	2.13	-1.27	-0.19	0.67	1.98
14.2	Was your chest tightness aggravated by daily activities?	2.69	-1.45	-0.47	0.25	1.55
14.3	Was your panting aggravated by daily activities?	2.43	-1.62	-0.58	0.12	1.50
14.4	Was your shortness of breath aggravated by daily activities?	2.52	-1.56	-0.54	0.19	1.54
15	Did you often feel upset?	2.20	-1.49	0.04	0.96	2.65
16	Did you often feel sad?	2.16	-1.16	0.30	1.46	2.67
17	Did you worry about your illness?	1.98	-1.33	-0.22	0.77	2.03
18.1	Was your cough aggravated by mood swings?	2.71	-0.81	0.23	0.92	2.03
18.2	Was your chest tightness aggravated by mood swings?	2.64	-0.98	0.06	0.82	2.17
18.3	Was your panting aggravated by mood swings?	2.63	-0.91	0.06	0.78	2.75
18.4	Was your shortness of breath aggravated by mood swings?	2.71	-0.87	0.11	0.86	2.73
19	Did the illness affect your contacts with friends, colleagues or neighbors?	2.27	-0.51	0.42	1.22	2.64
20	Was your illness affected by weather changes (cloudy, smoggy, damp, and sultry days, etc.)?	1.24	-3.08	-1.62	-0.40	1.75
21	Did your illness get worse when in a closed or poorly ventilated environment?	1.68	-1.84	-0.79	0.41	2.44

### Reliability

The SEM ranged from 0.266 to 0.300 using random method, while that ranged from 0.276 to 0.300 using MFI method (Table 3). On this basis, the calculated measurement reliability ( $r$ ) ranged from 0.910 to 0.929 using random method, while that ranged from 0.910 to 0.924 using MFI method (Table 3). The correlation coefficient of SEM and measurement reliability ( $r$ ) between the two methods were both 0.267 ( $P=0.040$ ;  $P=0.040$ ), and the independent sample  $t$ -test showed that there were no significant differences in SEM and measurement reliability ( $r$ ) between the two methods ( $t=-0.533$ ,  $P=0.594$ ;  $t=-0.472$ ,  $P=0.637$ ).

### Content validity

As was reported in our previous publication, the paper-and-pencil version of mCOPD-PRO had good content validity [9]. The CAT was developed based on the 27 items of the paper-and-pencil version of mCOPD-PRO, and therefore, was considered to have good content validity.

### Criterion validity

The theta ranged from  $-2.331$  to  $1.226$  using random method, while that ranged from  $-2.336$  to  $1.102$  using MFI method (Table 3). The correlation coefficient of theta between the two methods was 0.753 ( $P<0.001$ ), and the independent sample  $t$ -test showed that there were no significant differences in theta between the two methods ( $t=-0.514$ ,  $P=0.609$ ). The correlation coefficient between theta and COPD assessment test and mMRC scores using random method was 0.628 and 0.540 ( $P<0.001$ ;  $P<0.001$ ) respectively, while that using MFI method was 0.347 and 0.328 ( $P=0.007$ ;  $P=0.010$ ) respectively.

### Comparisons between the CAT and the paper-and-pencil version of mCOPD-PRO

The initial item of the paper-and-pencil version of mCOPD-PRO was all the same ("Did you cough?"), while that of the CAT was random using random method, and was all the same ("Was your chest tightness aggravated by daily activities?") using MFI method. All (27 items) were tested for the paper-and-pencil version, while about 11 items (reducing by 59.3%) were tested on average for the CAT using random method with at least eight

**Table 2** The average SEM and number of items administered, and correlation coefficient between simulated and true theta estimates in different simulation tests

Sample size	Average number of items administered	Average SEM	Correlation coefficient between simulated and true theta estimates
<i>Random method</i>			
60	10	0.290	0.975
100	10	0.290	0.976
300	10	0.291	0.976
500	10	0.291	0.970
1000	10	0.291	0.973
3000	10	0.291	0.972
5000	10	0.290	0.973
<i>MFI method</i>			
60	7	0.289	0.968
100	7	0.292	0.979
300	7	0.289	0.975
500	7	0.290	0.973
1000	7	0.289	0.974
3000	7	0.290	0.977
5000	7	0.290	0.977

*Abbreviations* SEM, Standard Error of Measurement; MFI, Maximum Fisher Information

items (four cases) and at most 26 items (only one case), and about seven items (reducing by 74.1%) on average using MFI method with at least six items (36 cases) and at most 20 items (only one case) (Additional file 10). The correlation coefficient between theta of the CAT and the paper-and-pencil version of mCOPD-PRO total scores using random and MFI methods was 0.919 and 0.760 ( $P < 0.001$ ;  $P < 0.001$ ) respectively. The correlation coefficient between theta and physiological, psychological and surrounding domain scores using random and MFI methods was presented in Table 4.

## Discussion

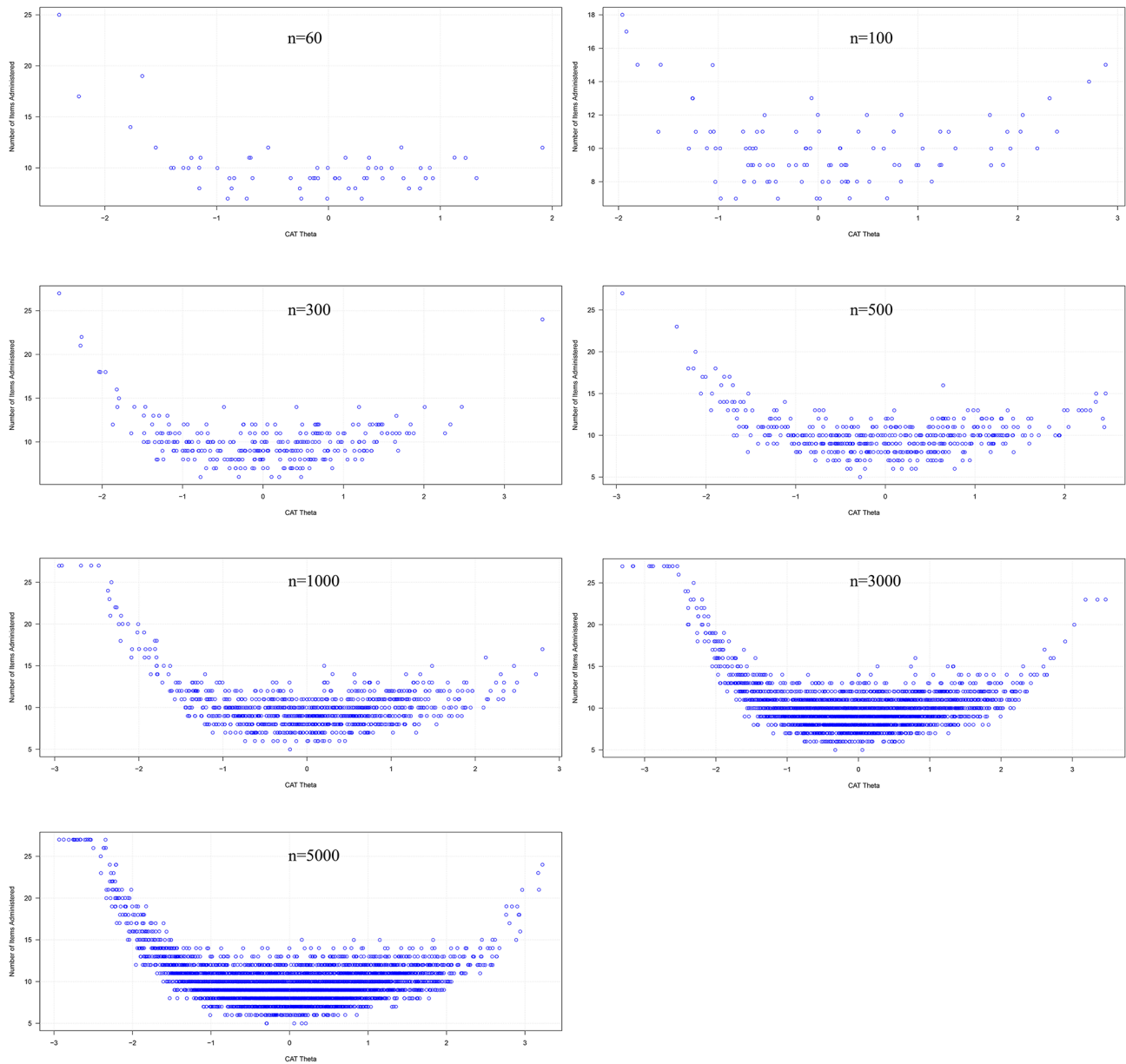
To our knowledge, this was the first study to develop and validate a comprehensive disease-specific CAT for PRO measurement of COPD by using the *Concerto* platform. Our study showed that the CAT is efficient, accurate, and user-friendly with good reliability and validity. Compared with the paper-and-pencil version of mCOPD-PRO, the average number of items tested for the CAT reduced by more than 50%, which indicated a significant reduction of respond burden. Moreover, the measurement accuracy was quite high. These findings are believed to contribute to the field.

As an open-source online adaptive testing platform, the *Concerto* platform has been used for neuropsychological testing and PRO measurement, and has been considered capable of harnessing the power of CAT and machine learning for developing and administering

advanced PRO measurement [23, 32–34]. Therefore, our study chose this platform to establish the CAT. The item bank of high quality is the basis to ensure the scientificity and maximize the advantages of CAT. In the current study, the item discrimination (a) and difficulty (b) were calibrated using IRT analysis. Our results suggested that the item parameters were generally ideal, which ensured the robustness of the CAT. As is well known, the sample size for the validation of CAT is important. However, it is difficult to determine the appropriate sample size by repeated clinical tests. Therefore, the CAT simulation with different sample sizes was performed based on the true participants' responses data of the paper-and-pencil version of mCOPD-PRO in the phase of validation. It was showed that the results of simulation tests were stable and reliable, and the sample size of 60 cases was enough. Therefore, the true participants' responses data from 60 participants was used for analysis in this study.

At present, the researches on psychometric properties of CAT for PRO instruments are lacking. In our study, the reliability and validity were evaluated by referring to relevant researches in the field of psychological and educational measurements. The measurement reliability ( $r$ ) was calculated through the SEM obtained from the CAT. As a result, the measurement reliability ( $r$ ) using random and MFI methods were both greater than 0.9, which indicated a good reliability of the CAT. As for criterion validity, the selection of standard instruments is quite important. Both COPD assessment test and mMRC were selected as our gold standards, which were recommended by the *Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease* [1, 30, 31]. The test results of the CAT showed moderate correlation with the two standard instruments using random method, and this indicated a good criterion validity. By comparison, the test results showed weak correlation with the two standard instruments using MFI method. This may be due to the concentration of items in psychological domain, and the results need to be further verified.

Although the item bank of the CAT and the paper-and-pencil version of mCOPD-PRO was the same, the order and number of items used were different. The initial item of the paper-and-pencil version was the first item "Did you cough?", while that of the CAT was random or the same "Was your chest tightness aggravated by daily activities?". A number of studies have assessed the effect of CAT on the length and accuracy of PRO measurement, robustly demonstrating that CAT can reduce the length of measurement by more than 50% while keeping excellent agreement between fixed-length measurement and CAT [34]. A CAT simulation study showed that when the stopping rule was matched to the reliability of published World Health Organization Quality of Life (WHOQOL)

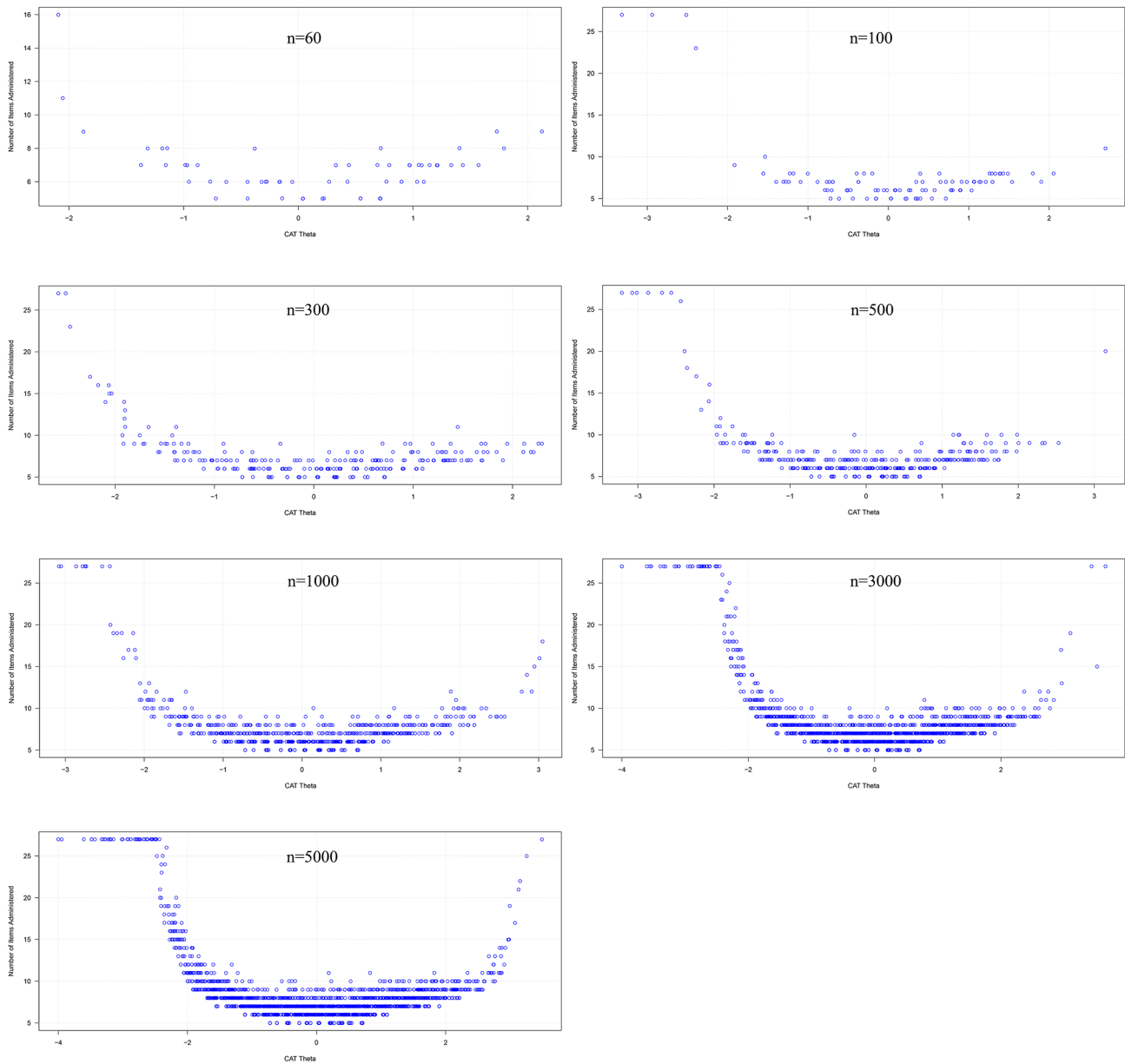


**Fig. 1** Number of items administered using random method in different simulation tests

assessment instrument, the item bank produced a measurement that was as reliable as the paper-and-pencil version of WHOQOL-BREF and WHOQOL-100 with 43% and 75% fewer items, respectively [35]. Our study showed that the average number of items tested for the CAT was about 11 using random method and seven using MFI method, respectively, which were almost consistent with our simulation tests. More importantly, compared with the paper-and-pencil version mCOPD-PRO, the average number of items tested for the CAT reduced by 59.3% and 74.1%, respectively, and the findings are similar to the results of the CAT simulation study mentioned above [35]. Accordingly, it is obvious that the respond burden of

the mCOPD-PRO was significantly reduced. That is, the CAT can improve the testing efficiency without reducing the measurement accuracy. In addition, the test results of the CAT showed strong correlation with the paper-and-pencil version of mCOPD-PRO, which supported the reliability of the CAT, and was in line with the published literature [34]. What’s more, the strong correlation of the test results between the two methods indicated the robustness of our CAT model.

There were limitations to our study. First of all, the CAT was developed and validated based on the *Concerto* platform, there may be limitation. Second, due to the different scoring rules and dimensions between the



**Fig. 2** Number of items administered using maximum Fisher information method in different simulation tests

CAT and the paper-and-pencil version of mCOPD-PRO, this study did not directly compare the differences of test results between the two versions. Third, given no consensus regarding the psychometric properties of CAT for PRO instruments, further studies are needed to verify our discoveries. In addition, the tests based on different item selection methods, ability estimation methods, or stopping criteria remain to be explored. Furthermore, the accuracy of parameter estimation would benefit from the survey with larger sample sizes in the future.

**Conclusions**

In conclusion, the comprehensive disease-specific CAT for PRO measurement of COPD is well developed, and has good reliability, content validity, and criterion validity. Compared with the paper-and-pencil version of mCOPD-PRO, the CAT can provide an efficient, accurate, and user-friendly measurement for PRO of COPD. Further studies remain to be explored in the future.



**Table 3** The SEM and measurement reliability (r) using random and MFI methods

Test No.	Random method		MFI method		Test No.	Random method		MFI method				
	Theta	SEM	r	Theta		SEM	r	Theta	SEM	r		
01	0.845	0.288	0.917	0.131	0.289	0.916	-1.029	0.291	0.915	-1.384	0.295	0.913
02	-0.260	0.300	0.910	-1.114	0.280	0.922	-1.494	0.291	0.915	-1.131	0.296	0.915
03	-0.491	0.290	0.916	-1.063	0.292	0.915	0.060	0.293	0.914	-0.019	0.291	0.915
04	-1.191	0.296	0.912	-1.320	0.295	0.913	0.894	0.287	0.918	0.251	0.290	0.916
05	0.014	0.291	0.915	0.280	0.290	0.916	-0.331	0.282	0.920	-0.280	0.296	0.912
06	-0.427	0.294	0.914	-0.348	0.297	0.912	-0.919	0.290	0.916	-0.781	0.298	0.911
07	-0.507	0.295	0.913	-0.791	0.298	0.911	-0.086	0.299	0.911	-0.006	0.291	0.915
08	1.226	0.299	0.911	1.028	0.276	0.924	0.469	0.295	0.913	0.298	0.290	0.916
09	-0.957	0.296	0.912	-1.227	0.289	0.916	-1.679	0.286	0.918	-1.402	0.288	0.917
10	0.100	0.298	0.911	-0.529	0.297	0.912	0.703	0.281	0.921	-0.662	0.280	0.922
11	-1.094	0.300	0.910	-0.621	0.297	0.912	-0.492	0.300	0.910	-0.291	0.296	0.912
12	-1.280	0.279	0.922	-0.088	0.292	0.915	0.369	0.299	0.911	0.045	0.290	0.916
13	0.319	0.283	0.920	0.501	0.288	0.917	0.742	0.297	0.912	1.102	0.291	0.915
14	-0.169	0.284	0.919	-0.937	0.285	0.919	-0.361	0.280	0.922	-0.044	0.291	0.915
15	0.100	0.294	0.914	-0.701	0.280	0.922	-0.372	0.284	0.919	-0.108	0.293	0.914
16	-1.314	0.296	0.912	-1.187	0.297	0.912	-0.280	0.294	0.914	0.270	0.290	0.916
17	-0.353	0.289	0.916	0.616	0.295	0.913	0.505	0.289	0.916	0.021	0.290	0.916
18	0.407	0.297	0.912	0.446	0.292	0.915	0.524	0.282	0.920	-0.701	0.280	0.922
19	-1.113	0.297	0.912	-0.591	0.297	0.912	-0.447	0.285	0.919	0.295	0.290	0.916
20	0.096	0.286	0.918	0.327	0.291	0.915	-1.421	0.294	0.914	0.386	0.291	0.915
21	0.073	0.295	0.913	0.280	0.290	0.916	-0.058	0.276	0.924	0.247	0.291	0.915
22	-0.994	0.294	0.914	-0.797	0.281	0.921	0.054	0.286	0.918	0.391	0.291	0.915
23	-0.941	0.293	0.914	-0.320	0.296	0.912	0.388	0.286	0.918	0.534	0.293	0.914
24	-0.237	0.294	0.914	0.083	0.290	0.916	-0.004	0.288	0.917	-0.701	0.280	0.922
25	-0.279	0.281	0.921	0.115	0.289	0.916	-1.873	0.295	0.913	-1.848	0.297	0.912
26	-0.162	0.299	0.911	-0.036	0.291	0.915	-2.097	0.299	0.911	-1.830	0.295	0.913
27	-0.742	0.290	0.916	-0.483	0.297	0.912	-0.439	0.295	0.913	-0.307	0.296	0.912
28	-2.331	0.298	0.911	-2.336	0.300	0.910	-1.941	0.266	0.929	-2.135	0.292	0.915
29	-0.990	0.293	0.914	-1.227	0.288	0.917	-0.919	0.296	0.912	-0.778	0.298	0.911
30	-0.678	0.296	0.912	-0.450	0.297	0.912	-0.814	0.283	0.920	-0.602	0.291	0.915

Abbreviations: SEM, Standard Error of Measurement; MFI, Maximum Fisher Information

**Table 4** The correlation coefficient between theta of the CAT and domain scores of the paper-and-pencil version of mCOPD-PRO

	mCOPD-PRO total scores	Physiological domain scores	Psychological domain scores	Surrounding domain scores
Random method	0.919	0.838	0.715	0.623
MFI method	0.760	0.569	0.864	0.535

Notes Data are expressed as correlation coefficient; all *P*-values < 0.001

Abbreviations CAT, Computerized Adaptive Testing; mCOPD-PRO, Modified Patient-reported Outcome Scale for Chronic Obstructive Pulmonary Disease; MFI, Maximum Fisher Information

#### Abbreviations

COPD	Chronic Obstructive Pulmonary Disease
PRO	Patient-reported Outcome
mCOPD-PRO	Modified Patient-reported Outcome Scale for Chronic Obstructive Pulmonary Disease
CAT	Computerized Adaptive Testing
IRT	Item Response Theory
RMSEA	Root-mean-square Error of Approximate
MFI	Maximum Fisher Information
SEM	Standard Error of Measurement
mMRC	Modified Medical Research Council Dyspnea Scale
WHOQOL	World Health Organization Quality of Life

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12955-024-02291-6>.

Supplementary Material 1: **Additional file 1** The matrix plot of item characteristic curves of the modified patient-reported outcome scale for chronic obstructive pulmonary disease (mCOPD-PRO)

Supplementary Material 2: **Additional file 2** The test information and standard error of measurement of the modified patient-reported outcome scale for chronic obstructive pulmonary disease (mCOPD-PRO)

Supplementary Material 3: **Additional file 3** The computerized adaptive testing procedure of an example

Supplementary Material 4: **Additional file 4** The percent of item usage using random method in different simulation tests

Supplementary Material 5: **Additional file 5** The correlation between simulated and true theta estimates using random method in different simulation tests

Supplementary Material 6: **Additional file 6** The test information and standard error of measurement using random method in different simulation tests

Supplementary Material 7: **Additional file 7** The percent of item usage using maximum Fisher information method in different simulation tests

Supplementary Material 8: **Additional file 8** The correlation between simulated and true theta estimates using maximum Fisher information method in different simulation tests

Supplementary Material 9: **Additional file 9** The test information and standard error of measurement using maximum Fisher information method in different simulation tests

Supplementary Material 10: **Additional file 10** The number of items tested for the computerized adaptive testing

#### Acknowledgements

The authors thank all the experts and patients involved in this study for their kind help.

#### Author contributions

JJW was responsible for the modeling and analysis of data, and drafted the manuscript. YX and ZZ were responsible for the interpretation of data. YX and JSL conceived this study and revised the manuscript. All authors read and approved the final manuscript.

#### Funding

This study was supported by National Natural Science Foundation of China (81473648, 81830116), Qi-Huang Chief Scientist Project of National Administration of Traditional Chinese Medicine ([2020]219), and Central Plains Thousand People Program (ZYQR201810159).

#### Data availability

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

This study was approved by the Institution Review Board of the First Affiliated Hospital of Henan University of Chinese Medicine (2015HL-048). Written informed consent was obtained from all participants included in this study.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Collaborative Innovation Center for Chinese Medicine and Respiratory Diseases Co-constructed by Henan Province & Education Ministry of P.R. China, Henan University of Chinese Medicine, 156 Jinshui East Road, Zhengzhou 450046, China

<sup>2</sup>Henan Key Laboratory of Chinese Medicine for Respiratory Disease, Henan University of Chinese Medicine, 156 Jinshui East Road, Zhengzhou 450046, China

<sup>3</sup>Department of Respiratory Diseases, the First Affiliated Hospital of Henan University of Chinese Medicine, 19 Renmin Road, Zhengzhou 450003, China

Received: 3 February 2024 / Accepted: 27 August 2024

Published online: 04 September 2024

#### References

- Agusti A, Celli BR, Criner GJ, Halpin D, Anzueto A, Barnes P, et al. Global Initiative for Chronic Obstructive Lung Disease 2023 Report: GOLD Executive Summary. *Eur Respir J*. 2023;61:2300239.
- Al Wachami N, Guennouni M, Iderdar Y, Boumendil K, Arraji M, Mourajid Y, et al. Estimating the global prevalence of chronic obstructive pulmonary disease (COPD): a systematic review and meta-analysis. *BMC Public Health*. 2024;24:297.
- Wang C, Xu J, Yang L, Xu Y, Zhang X, Bai C, et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. *Lancet*. 2018;391:1706–17.
- World Health Organization. The top 10 causes of death. 2020. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>. Accessed 1 Feb 2024.
- Jones PW. Health status and the spiral of decline. *COPD*. 2009;6:59–63.
- U.S. Food and Drug Administration. Guidance for Industry Patient-reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. [www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims). Accessed 1 Feb 2024.
- Cazzola M, MacNee W, Martinez FJ, Rabe KF, Franciosi LG, Barnes PJ, et al. Outcomes for COPD pharmacological trials: from lung function to biomarkers. *Eur Respir J*. 2008;31:416–69.

8. Weldam SW, Schuurmans MJ, Liu R, Lammers JW. Evaluation of quality of Life instruments for use in COPD care and research: a systematic review. *Int J Nurs Stud*. 2013;50:688–707.
9. Li J, Wang J, Xie Y, Feng Z. Development and validation of the modified patient-reported Outcome Scale for Chronic Obstructive Pulmonary Disease (mCOPD-PRO). *Int J Chron Obstruct Pulmon Dis*. 2020;15:661–9.
10. Tan Q, Cai Y, Li Q, Zhang Y, Tu D. Development and validation of an Item Bank for Depression Screening in the Chinese Population using computer adaptive testing: a Simulation Study. *Front Psychol*. 2018;9:1225.
11. Banerjee S, Plummer O, Abboud JA, Deirmengian GK, Levicoff EA, Courtney PM. Accuracy and Validity of Computer Adaptive Testing for Outcome Assessment in patients undergoing total hip arthroplasty. *J Arthroplasty*. 2020;35:756–61.
12. Gibbons RD, Weiss DJ, Frank E, Kupfer D. Computerized adaptive diagnosis and testing of Mental Health disorders. *Annu Rev Clin Psychol*. 2016;12:83–104.
13. Rebollo P, Castejón I, Cuervo J, Villa G, García-Cueto E, Díaz-Cuervo H, et al. Validation of a computer-adaptive test to evaluate generic health-related quality of life. *Health Qual Life Outcomes*. 2010;8:147.
14. Norweg A, Ni P, Garshick E, O'Connor G, Wilke K, Jette AM. A multidimensional computer adaptive test approach to dyspnea assessment. *Arch Phys Med Rehabil*. 2011;92:1561–9.
15. Choi SW, Victorson DE, Yount S, Anton S, Cella D. Development of a conceptual framework and calibrated item banks to measure patient-reported dyspnea severity and related functional limitations. *Value Health*. 2011;14:291–306.
16. Yount SE, Choi SW, Victorson D, Ruo B, Cella D, Anton S, et al. Brief, valid measures of dyspnea and related functional limitations in chronic obstructive pulmonary disease (COPD). *Value Health*. 2011;14:307–15.
17. Yount SE, Atwood C, Donohue J, Hays RD, Irwin D, Leidy NK, et al. Responsiveness of PROMIS® to change in chronic obstructive pulmonary disease. *J Patient Rep Outcomes*. 2019;3(1):65.
18. Paap MCS, Kroeze KA, Glas CAW, Terwee CB, van der Palen J, Veldkamp BP. Measuring patient-reported outcomes adaptively: multidimensionality matters! *Appl Psychol Meas*. 2018;42:327–42.
19. Paap MCS, Kroeze KA, Terwee CB, van der Palen J, Veldkamp BP. Item usage in a multidimensional computerized adaptive test (MCAT) measuring health-related quality of life. *Qual Life Res*. 2017;26:2909–18.
20. O'Hoski S, Richardson J, Kuspinar A, Wald J, Goldstein R, Beauchamp MK. A brief measure of life participation for people with COPD: validation of the computer adaptive test version of the late life disability instrument. *COPD*. 2021;18:385–92.
21. O'Hoski S, Kuspinar A, Richardson J, Wald J, Goldstein R, Beauchamp MK. Responsiveness of the late life disability instrument to pulmonary rehabilitation in people with COPD. *Respir Med*. 2023;207:107113.
22. Ho EH, Verkuilen J, Fischer F. Measuring individual true change with PROMIS using IRT-based plausible values. *Qual Life Res*. 2023;32:1369–79.
23. Scalise K, Allen DD. Use of open-source software for adaptive measurement: Concerto as an R-based computer adaptive development and delivery platform. *Br J Math Stat Psychol*. 2015;68:478–96.
24. Tu DB, Zheng CJ, Dai BY, Wang WY. Computerized adaptive testing: theory and method. Beijing: Beijing Normal University; 2017.
25. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, 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:S22–31.
26. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1–55.
27. Wen ZL, Hau KT, Marsh HW. Structural equation model testing: cutoff criteria for goodness of fit indices and chi-square test. *Acta Psychol Sin*. 2004;36:186–94.
28. Chen WH, Thissen D. Local dependence indexes for item pairs using item response theory. *Educational Behav Stat*. 1997;22:265–89.
29. Van der Ark LA. New developments in Mokken scale analysis in R. *J Stat Softw*. 2011;48:1–27.
30. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. *Eur Respir J*. 2009;34:648–54.
31. Perez T, Burgel PR, Paillasseur JL, Caillaud D, Deslée G, Chanez P, et al. Modified Medical Research Council scale vs baseline Dyspnea Index to evaluate dyspnea in chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*. 2015;10:1663–72.
32. Yudien MA, Moore TM, Port AM, Ruparel K, Gur RE, Gur RC. Development and public release of the Penn Reading Assessment Computerized Adaptive Test (PRA-CAT) for premorbid IQ. *Psychol Assess*. 2019;31:1168–73.
33. Harrison C, Trickett R, Wormald J, Dobbs T, Lis P, Popov V, et al. Remote Symptom Monitoring with Ecological momentary computerized adaptive testing: pilot cohort study of a platform for frequent, Low-Burden, and personalized patient-reported outcome measures. *J Med Internet Res*. 2023;25:e47179.
34. Harrison C, Loe BS, Lis P, Sidey-Gibbons C. Maximizing the potential of patient-reported assessments by using the Open-Source Concerto platform with computerized adaptive testing and machine learning. *J Med Internet Res*. 2020;22:e20950.
35. Gibbons C, Bower P, Lovell K, Valderas J, Skevington S. Electronic quality of Life Assessment using computer-adaptive testing. *J Med Internet Res*. 2016;18:e240.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.