



Mobile Online Computer-Adaptive Tests (CAT) for Gathering Patient Feedback in Pediatric Consultations

Tsair-Wei Chien^{1,2}, Wen-Pin Lai³, Ju-Hao Hsieh^{3,*}

¹Research Department, Chi-Mei Medical Center, Tainan, Taiwan

²Department of Hospital and Health Care Administration, Chia-Nan University of Pharmacy and Science, Tainan, Taiwan

³Department of Emergency Medicine, Chi-Mei Medical Center, Tainan, Taiwan

Email address:

wenpin.lai@msa.hinet.net (Wen-Pin Lai), smile@mail.chimei.org.tw (Ju-Hao Hsieh)

*Corresponding author

To cite this article:

Tsair-Wei Chien, Wen-Pin Lai, Ju-Hao Hsieh. Mobile Online Computer-Adaptive Tests (CAT) for Gathering Patient Feedback in Pediatric Consultations. *Applied and Computational Mathematics*. Special Issue: Some Novel Algorithms for Global Optimization and Relevant Subjects. Vol. 6, No. 4-1, 2017, pp. 64-71. doi: 10.11648/j.acm.s.2017060401.16

Received: December 19, 2017; **Accepted:** January 9, 2017; **Published:** February 6, 2017

Abstract: *Background:* Few studies have used online patient feedback from smartphones for computer adaptive testing (CAT). *Objective:* We developed a mobile online CAT survey procedure and evaluated whether it was more precise and efficient than traditional non-adaptive testing (NAT) when gathering patient feedback about their perceptions of interaction with a physician after a consultation. *Method:* Two hundred proxy participants (parents or guardians) were recruited to respond to twenty 5-point questions (the P4C_20 scale) about perceptions of doctor-patient and doctor-family interaction in clinical pediatric consultations. Through the parameters calibrated using a Rasch partial credit model (PCM) and a Rasch rating scale model (RSM), two paired comparisons of empirical and simulation data were administered to calculate and compare the efficiency and precision of CAT and NAT in terms of shorter item length and fewer counts of difference number ratio (< 5%) using independent *t* tests. An online CAT was designed using two modes of PCM and RSM for use in clinical settings. *Results:* The graphical online CAT for smartphones used by the parents or guardians of pediatric hospital patients was more efficient and no less precise than NAT. *Conclusions:* CAT-based administration of the P4C_20 substantially reduced respondent burden without compromising measurement precision.

Keywords: Computer Adaptive Testing, Non-daptive Testing, Partial Credit Model, Rasch Analysis, Rating Scale Model

1. Introduction

Two major methods are used to assess patient clinical outcomes and patient perceptions in clinical settings [1]: (a) a lengthy questionnaire and (b) a rapid short-form scale [2, 3]. Each has advantages and disadvantages. Both are traditional pencil-and-paper assessments with a large respondent burden because they require patients to answer questions that are sometimes too easy and sometimes too difficult and do not provide any additional information [4].

If a patient has a tendency toward a symptom (e.g., skin cancer, dengue fever, or a satisfaction perception level) that can be thought as a latent trait [5-7], we want to transform their observed scores on a unidimensional scale to unobserved characteristic attributes (i.e., the aforementioned symptom).

The item response theory (IRT)-based Rasch family models [8-11] have often been used to examine whether a scale is unidimensional and appropriate for assessing the symptom (e.g., a latent trait or, in this study, a perception of satisfaction). Computer adaptive testing (CAT) was also used to overcome the inefficient disadvantages of traditional pencil-and-paper non-adaptive testing (NAT) assessment [6, 7].

Few studies have used the more complicated Rasch partial credit model (PCM) [11] (polytomous but with a different number of categories for each item, in contrast to the dichotomous Rasch model [12, 23]) or the rating scale model (RSM) [10] (with the same number of categories across items) in clinical settings. Neither have many studies used mobile

online CAT assessment with PCM.

Patient-centered care that includes patient participation is widely accepted as a key aim of hospitals and healthcare systems [14]. Clinical consultations provided by physicians are an important aspect of the doctor's role [15] and are determiners of the overall quality of patient care. This is true in part because of the consumerist approach to healthcare, which [16] requires doctors to be more accountable to their patients [17-19], and because many hospitals are required by accreditation institutes to use questionnaires to assess patient and family satisfaction with physician performance as part of routine self-management [14, 20]. To improve the quality of medical practice, these questionnaires draw attention to issues such as the doctor's communication skills and attitudes [21, 22]. The assessment of individual and group performance by physicians has thus gained increasing prominence worldwide [23].

After gathering patient feedback about their perceptions of doctor-patient and doctor-family interaction, we wanted to (a) evaluate the difference person standard errors between CAT and NAT, (b) compare the precision and efficiency of CAT and NAT under different models of RSM and RSM within different scenarios using empirical and simulation data, and (c) design a mobile online CAT survey procedure to evaluate its precision and efficiency.

2. Methods

Data source

The study sample was recruited from pediatric patients who visited the emergency room (ER) of a 1300-bed medical center in Taiwan. During the last 14 days of November 2013, fifteen first-visit patients per day (3 each morning, afternoon, evening, night, and at midnight; total: 210) who had just finished a consultation with an ER pediatrician were enrolled based on the last number (3, 6, and 9) of their hospital chart, and then patient proxies (family members: parents, siblings, or other relatives) were asked to complete a questionnaire.

This study was approved and monitored by the Research and Ethical Review Board of the Chi-Mei Medical Center (No.CMFIRB098231).

Instrument

The 20 items of the patient-centered participation care in consultations involving children (refer to the P4C_20 scale) were selected from the literature [15, 24-26] and revised by a consensus panel of 12 members (7 ER pediatricians and 5 ER nurses). Each item was assessed using a 5-point Likert scale with a range from 1 (very disagree) to 5 (very agree). In keeping with good practice in item selection, the questionnaires were tested in 30 iterative pilot trials to ensure that the question expression, the rating scales, and the layout were comprehensive, and acceptable to respondents.

Data Analysis

Rasch Winsteps software [27] was used to estimate item and person parameters of the sample under two Rasch family models of PCM (choosing eligible categories endorsed by at least one respondent) and RSM (appropriate for 5 points

across all items), respectively. Two paired comparisons of empirical and simulation datasets were made (Figure 1) to calculate the precision and efficiency of CAT and NAT in terms of shorter item length and fewer different count ratios (< 5%) using independent *t* tests [28] under different models of RSM and of RSM within different scenarios using empirical and simulation data. We examined whether those 20 items fit the Rasch unidimensional measurement requirement using the 3-step detection approaches [12, 30]: parallel analysis [31], Infit and Outfit mean square errors < 1.5 [29], and the Rasch PCA of residuals. With item parameters (i.e., overall and step difficulties), an online CAT was designed with two modes of PCM and RSM for use in clinical settings.

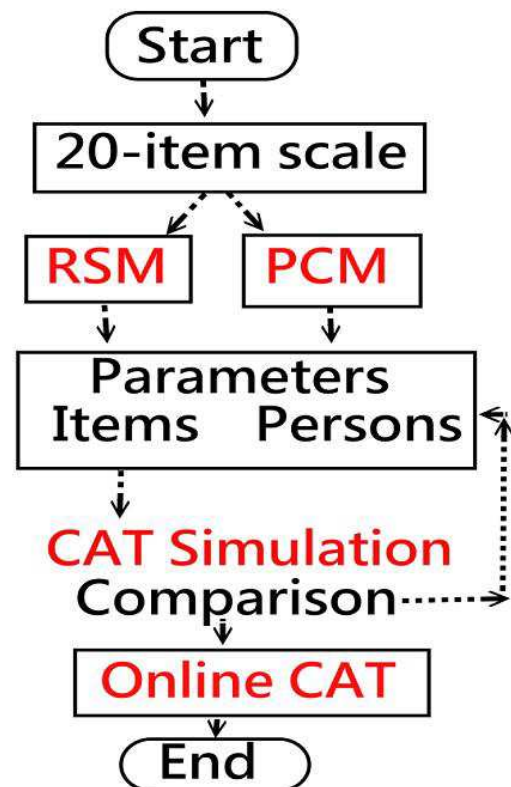


Figure 1. Study simulation and CAT flowchart.

Mobile online CAT designed for smartphones

The initial item was randomly selected from 20 items once the CAT was begun. The provisional person measure is estimated by the maximum likelihood estimation (MLE) [32] using an iterative Newton-Raphson procedure [33]. The final measure is determined by the maximum of the log-likelihood function before terminating the CAT. The next item selection is based on the highest Fisher information (i.e., item variance) of the remaining unanswered items interacted with the provisional person measure [33]. The results (theta, standard error [SE], Infit, and Outfit yielded by the author-made module) are equivalent to the Winsteps estimation.

The three-item termination rule set in the CAT module is (a) the person estimated SE standard error ($SE = 1/\sqrt{\Sigma \text{variance (i)}}$), where (i) refers to the CAT finished items responded to by a person [13]) less than 0.40 (equivalent to the

test Cronbach's $\alpha = 0.84 = [1 - SE^2]$ according to the formula: $MSE = SD \times \sqrt{(1 - \text{reliability})}$ [6, 7, 34] and the average person estimated SE yielded by the empirical study sample). The minimum number of items required for completion was 7 to meet the minimal individual person reliability. The designed CAT will be terminated (b) once person outfit MNSQ is > 20 (mainly due to aberrant responses endorsed by the examinee) or (c) the last 5 average consecutive person estimation changes are < 0.05 after the completed items reach the minimal requirement of 7.

Simulation to verify the advantage of CAT over NAT

To compare different CAT effects on an empirical dataset, 1,000 normally distributed respondents (Mean = 0, SD = 1) were manipulated and simulated [35] under the Rasch PCM and RSM models. Comparisons of person mean, SEs, item length, person measure correlation coefficients, and different number ratios between CAT and NAT across all scenarios (2 models, 2 datasets) was made to determine whether CAT has the advantage over NAT. We ran an author-made Visual Basic for Applications (VBA) module in Microsoft Excel to conduct the simulation study (<http://youtu.be/W-EOJdW8oXE>) and demonstrated an online CAT assessment used for smartphones.

3. Results

Participants

Of the 210 potential participants recruited, 10 were excluded because of response errors and missing values on the questionnaire. The demographic characteristics (gender, age, relationship to child, and education) of the proxies of the ER pediatric sample show that the most frequent accompanying adult was the child's mother (80.5%), most (65.5%) of whom were between 31 and 40 years old (Table 1).

Table 1. Distribution of proxy characteristics.

Proxy Characteristics	Count	%
Gender		
Male	31	15.5
Female	169	84.5
Age (years)		
Under 30	55	27.5
31-40	131	65.5
41-50	13	6.5
51-60	1	0.5
Relationship		
Father	31	15.5
Mother	161	80.5
Grandparent	1	0.5
Sibling	3	1.5
Babysitter	1	0.5
Others	3	1.5
Education		
Less than high school	5	2.5
High school graduate	84	42.0
Some college	50	25.0
Bachelor's degree	52	26.0
Post-graduate	9	4.5

The unidimensional P4C_20 scale

The P4C_20 scale can be considered unidimensional: (1) parallel analysis was used to extract one factor [31] and had an acceptable dimension coefficient (DC) (> 0.70) [36]; (2) all infit and outfit mean squares for the 20 items were in a range of 0.5 to 1.5; (3) the Rasch residual DC calculated using PCA eigenvalues was too small (< 0.60) to have another domain component in the scale.

All of the reliabilities (including Rasch reliability and Cronbach's σ) were > 0.80 (Table 2, row 1), and all of the DCs and residual DCs were > 0.70 and < 0.60 (Table 2, row 1). The overall item difficulties and step difficulties for PCM and RSM (see threshold difficulties beneath Table 3) were then applied to simulate study data [34].

Table 2. Efficiency and precision of CAT compared with NAT

		Empirical Data				Simulation Data			
No	Type	RSM		PCM		RSM		PCM	
	A	B	C	D	E	F	G	H	I
1	Scale properties: NAT (20 items)	0.80 ^a DCr ^c	0.82 ^b DC ^d	0.80 ^a DCr ^c	0.82 ^b DC ^d	0.89 ^a DCr ^c	0.90 ^b DC ^d	0.81 ^a DCr ^c	0.89 ^b DC ^d
2	$\leq 0.6 \cap \geq 0.70$	0.56	0.71	0.55	0.71	0.49	0.87	0.51	0.86
	Study goals: Standard error	mean	SE	mean	SE	mean	SE	mean	SE
3	NAT	1.04	0.38	0.57	0.39	0.00	0.34	−0.79	0.27
4	CAT	0.52	0.43	−0.26	0.43	−0.27	0.42	−0.04	0.42
	Efficiency	length	saving	length	saving	length	saving	length	saving
5	CAT	10.55 ^e	47.25% ^f	11.86 ^e	40.70% ^f	9.55 ^e	52.25% ^f	10.28 ^e	48.60% ^f
	Precision	diff. ratio	Corr	diff. ratio	Corr	diff. ratio	Corr	diff. ratio	Corr
6	CAT	0.30% ^g	0.88 ^h	0.50% ^g	0.89 ^h	0.30% ^g	0.96 ^h	0.40% ^g	0.97 ^h

SE = standard error of the mean.

^a Rasch rel = Rasch person reliability; ^b Alpha = Cronbach's σ ; ^c DCr = Dimension coefficient of Rasch residual; ^d DC = Dimension coefficient

^e CIL = Average CAT item length; ^f % = $1 - \text{CIL}/20$; ^g Diff (%) = Different number ratio compared with the 20-item data set; ^h Corr = Correlation coefficient of person theta to NAT.

Table 3. Rasch analysis of the 20 study items.

Content	Difficulty		Threshold Difficulty			
	RSM ^a	PCM	1	2	3	4
1. I understood all of the doctor's explanations	-0.73	-0.42	-0.58	-0.9	-1.2	2.68
2. I feel that the doctor used too much medical jargon.*	0.32	0.83	-1.3	-0.74	2.04	
3. I feel confident about the doctor's professional knowledge.	0.18	-0.17	-2.74	-0.49	-0.06	3.29
4. The doctor repeatedly answered my questions about my child's illness when I misunderstood.	-0.45	0.18	-0.97	-1.1	2.07	
5. I feel that the doctor explained the prescription and treatment in sufficient detail.	-0.21	-0.64	-3.06	0.45	-0.18	2.79
6. I feel the doctor gave us an appropriate amount of consultation time.	-0.25	0.25	-1.92	-0.74	2.66	
7. The consultation time was too short to communicate with doctor.*	0.25	0.83	-2.26	-0.59	2.85	
8. The doctor was considerate and friendly enough.	2.41	0.8	-2.15	0.82	1.33	
9. The doctor always encouraged me to describe my child's illness.	-0.81	-0.21	-1.25	-1.18	2.42	
10. The doctor often used Yes/No dichotomy questions when asking about my child's illness	-0.26	0.45	-1.7	-1.3	3	
11. The doctor listened to and was concerned about my description of my child's illness.	-1.12	-0.96	-0.58	-1.5	-0.84	2.92
12. The doctor immediately responded to my questions about my child's illness.	-1.61	-1.25	-1.84	-0.66	2.5	
13. The doctor seldom made eye contact with us when in consultation.*	0.24	0.05	-1.44	-0.42	-0.21	2.07
14. The doctor made conclusions after the consultation.	-0.66	-0.83	-2.24	-0.33	0.05	2.52
15. The doctor was gentle with and sympathetic to my child.	-1.09	-1	0.2	-2.72	0.03	2.5
16. The doctor talked to my child if necessary.	-0.95	-0.91	-1.23	-1.19	-0.35	2.77
17. The doctor described how to use drugs in sufficient detail.	3.93	1.94	-1.97	1.97		
18. The doctor told me the side effects of drugs.	3.54	2.54	-2.74	0.04	2.7	
19. The doctor increased my confidence after the consultation.	-1.56	-0.45	-2.01	2.01		
20. The doctor told me the risk symptoms for my child's illness.	-1.18	-1.01	-2.16	-0.9	3.05	

*Inverse scoring; ^a RSM fixed threshold difficulties: -3.19, -0.27, 0.12, 3.34

Comparing the advantages of CAT and NAT

About person SE

Because CAT items were shorter than NAT items, NAT had a slightly smaller SE than did CAT (Table 2, rows 3 and 4). Simulation data with a higher tendency toward unidimensionality (see higher DC and lower DCr in Table 2, row 2) had a smaller SE than did empirical data in both the NAT and CAT scenarios; moreover, item lengths were shorter (Table 2, row 5). In the CAT scenario, SE differences between the PCM and RSM models were not significant.

About efficiency and precision

The simulation data (with a higher tendency toward unidimensionality) were more highly correlated, precise, and efficient than were the empirical data (Table 2, rows 5 and 6) in both CAT and NAT scenarios. There were no significant differences in correlation between models, and only slight differences in efficiency and precision. The differences in all of the number ratios compared with the 20-item data set were less than 5% (Table 2, row 6), which indicated that CAT was substantially more efficient than was NAT without compromising the precision of assessment between models or between datasets (Table 2, row 5).

A mobile online CAT module designed for smartphones

We developed a mobile CAT survey procedure (Figure 2, QR-code) to demonstrate the CAT application for two models. The item-by-item CAT process is shown in Figure 2. Person fit statistics (MNSQ of Infit and Outfit) depict normal and aberrant respondent behaviors. Person theta is the provisional ability estimated by the CAT module. The MSE is the person SE generated by the formula $1/\sqrt{(\sum \text{variance}(i))}$, where (i) refers to the CAT items responded to by a person [13]. In addition, the Rasch residual (resi) is the average of the last 5 change differences between the pre-and-post estimated abilities on each CAT step. CAT will stop if the value of the resi is < 0.05 . The correlation coefficient of person theta to NAT (corr) refers to the correlation coefficient between the CAT estimated measures and its step series numbers using the last 5 estimated theta values, which shows whether the final estimation theta trend is positively or negatively convergent. The flatter the theta trend means are, the higher is the probability of the person measure being convergent to a final estimation. More items will result in a lower SE. A $|Z|$ score > 2.0 denotes an unexpected response interaction between the final person measure and the respective item difficulty.

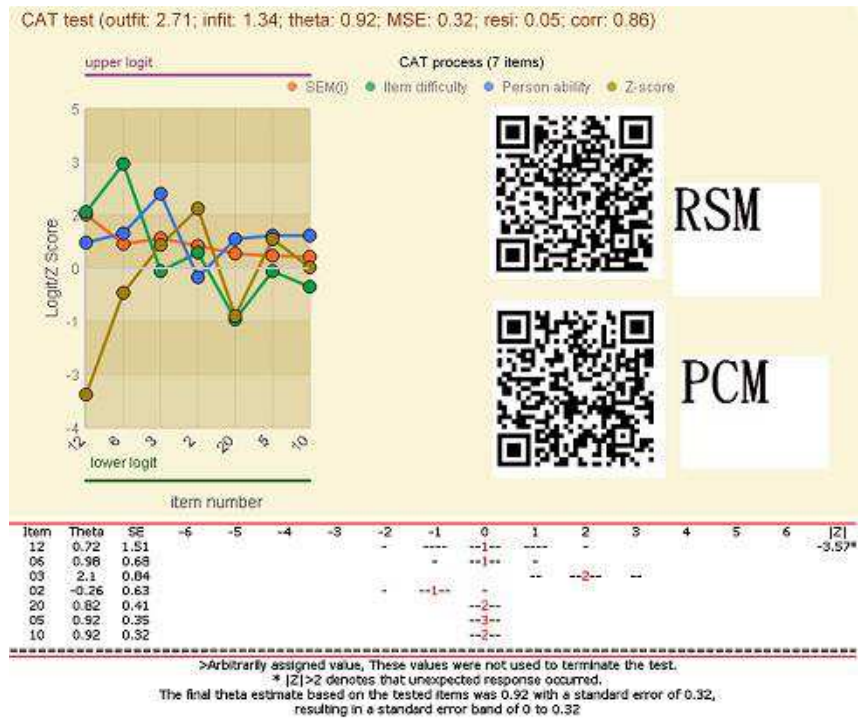


Figure 2. A graphical CAT report shown after each response.

4. Discussion

Key finding

We found that the simulation data yielded a smaller person SE, a higher correlation with the NAT person measure (more precise), and a shorter CAT item length (more efficient). Differences in smaller person SE, efficiency, and precision between the PCM and RSM models was nonsignificant. CAT had a slightly wider SE than did NAT because of a longer item length; however, efficiency reduced the respondent response burdens along with equivalent measures with NAT. A mobile online CAT for gathering patient feedback on the doctor is feasible on smartphones. Interested readers can practice it by using the QR-code shown in Figure 2.

What this adds to what is known

We confirmed that the patient perception of communication and interaction with the doctor is practicable, workable, and viable on smartphones whether using PCM or RSM. However, the greater tendency toward unidimensionality of a scale (like the simulation data in the current study) yields significantly more efficiency and precision than does NAT, dependent upon the extent of unidimensionality with higher DC and lower DCr.

Eastaugh [39] said that “nations with global budgets have better health statistics, and lower costs, compared to the United States. With global budgets, these countries employ 75 to 85% fewer employees in administration and regulation, but patient satisfaction is almost double the rate in the United States”, which indicates that gathering feedback from patients is essential improving the quality of care, especially in an age of patient involvement and patient-centered participation in healthcare.

What are the implications? What should be changed?

Our findings are consistent with the literature[37, 40-42], and they support the notion that CAT is more efficient than NAT. Patient needs and communication challenges vary greatly in different clinical settings [43] because different types of patients have their own characteristics and requirements [15]. For instance, an ER physician consulting the families of pediatric patients is in a different situation than when consulting the families of adults [15, 44]. It is necessary to illustrate the example (such as items in the P4C_20 scale) using a scientific approach with pediatricians to answer the following research questions: *What kind of checklist can we create to mitigate doctor consultation behavior errors?* [45] and *What kind of performance indicators can we set to continuously improve the quality of patient-centered care?* [15, 43]. The smartphone feedback from the proxies of pediatric patients after a consultation should be useful for hospitals and clinics because it will tell them, in the patients’ family members’ own words, what the latter want to know in consultations with their children’s doctors.

Using the mobile online CAT module to efficiently and precisely gather responses from patients is feasible and practical. Outfit MNSQ values ≥ 2.0 can be used to examine whether patient responses are distorted or abnormal, which means that many more responses unexpectedly fit the model’s criteria and were deemed careless, mistaken, cheating, or awkward [6, 7, 43] (e.g., Outfit MNSQ 2.71; the most unexpected responses show an asterisk (*) on the |Z| column in Figure 2 if $|Z| > 2.0$). Another advantage of IRT over traditional classic test theory [33, 46] is that it provides more information.

In addition, the graphical representations in Figure 2

IRT-based CAT users know that any significantly aberrant or cheating behavior on CAT will be detected by the module algorithm.

Strengths of this study

We confirmed that CAT has the advantages of both forms: precision and efficiency. In addition, this paper used the Rasch PCM (instead of the dichotomy or RSM models used in other studies [6, 7, 12, 13]) to design a CAT smartphone app and used it to assess pediatric patient proxies about the quality of their consultations with ER pediatricians, which has not been done before. We also considered two situations that never discussed in previous CAT studies: (1) the inversed items (like items 2, 7, and 13 in Table 3) are automatically reversed for estimating measures in the CAT process; (2) all the unanswered items will be automatically filled in an appropriate response in compliance with Rasch model's requirement [35].

Furthermore, it is also easy to set up any form of online CAT assessment only if the app designer uploads relevant parameters into the database, such as the type of IRT model; threshold difficulties; the number of questions in the item bank, test or questionnaire; and whether to show plots; etc.

We simulated data using models with different item lengths to execute CAT (<http://youtu.be/W-EOJdW8oXE> for more detail). Interested readers are encouraged to request the Excel-type module.

As with all forms of Web-based technology, advances in mobile health and health communication technology are rapidly increasing [47]. The online CAT for smartphones is promising and worth promoting.

Limitations

This study has some limitations. First, no detailed or comprehensive examination (e.g., DIF [48]) was done for the item parameters invariant across groups; hence, our findings are probably not generalizable. Second, our online versions of the CAT app are in English and Chinese only and need to be translated into other languages. Third, the CAT graphic shown in Figure 2 might be confusing and difficult for patients and their family members to interpret. An option to close this window and replace it with a simpler visualization, if the user prefers, is necessary.

Applications

Our online CAT smartphone app for gathering feedback on doctor consultation is feasible. CAT designers might want to expand or otherwise modify the item pool, or replace the items so that the app can be used for other kinds of information gathering.

It is necessary to point out that the (1) item overall (i.e., on average) and step (threshold) difficulties of the items must be calibrated in advance using Rasch analysis (as in Table 3); (2) pictures used for the subject or response categories for each question should be well-prepared with a web link that can be shown simultaneously with the item that appears in the CAT animation; (3) the app can be adapted for use with many kinds of IRT-based models, such as the more complicated generalized partial credit model CAT with a discrimination parameter for each item. Moreover, the correct parameters

corresponding to the exact fields of the database need to be uploaded; (4) the QR-code can be pasted onto the individual patient's receipt, prescription, or outside the room for easy access to respond with personal perceptions about consultations with the doctor; and (5) the final measure (say, 0.93 logits) can be shown with a T-score (mean = 50, SD = 10) at 59.30 for easy interpretation by the public. Multimedia Appendix 1 (or see http://youtu.be/brl_E_E9124) demonstrates the CAT app that can be viewed and practiced online by interested readers.

5. Conclusion

The CAT for P4C_20 scale reduced the respondents' burden without compromising measurement precision and increased response efficiency. The CAT app used for smartphones is recommended for the online assessment of other kinds of information from patients in future.

Abbreviations

CAT: computer adaptive testing
DIF: differential item functioning
IRT: item response modeling
NAT: non-adaptive testing
MSE: standard error of measurement
PCA: principal component analysis
PCM: partial credit model
RSM: Rasch rating scale model
VBA: Visual Basic for Applications

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors have read and approved the final manuscript. TW developed the study concept and design. TW and WP analyzed and interpreted the data. TW and JH drafted the manuscript and all authors have provided critical revisions for important intellectual content. The study was supervised by TW.

Acknowledgement

This study was supported by grant CMFIRB098231 from the Chi Mei Medical Centre, Taiwan. We are grateful to Ching-Chin Huang, Fu-Mei Dai and Huang-Lan Li, members of the Chi-Mei Cancer Center, for their invaluable administrative assistance and the collection of data.

References

- [1] Eack SM, Singer JB, Greeno CG. Screening for anxiety and depression in community mental health: the Beck Anxiety and Depression Inventories. *Community Ment Health J* 2008;44 (6): 465-474.

- [2] Ramirez Basco M, Bostic JQ, Davies D, et al. Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry* 2000;157 (10): 1599-1605.
- [3] Shear MK, Greeno C, Kang J, et al. Diagnosis of nonpsychotic patients in community clinics. *Am J Psychiatr* 2000;157: 581-587.
- [4] De Beurs DP, de Vries AL, de Groot MH, de Keijser J, Kerkhof AJ. Applying computer adaptive testing to optimize online assessment of suicidal behavior: a simulation study. *J Med Internet Res* 2014;16 (9): e207.
- [5] Lai WP, Chien TW, Lin HJ, Su SB, Chang CH. A screening tool for dengue fever in children. *Pediatr Infect Dis J* 2013;32 (4): 320-324.
- [6] Chien TW, Wang WC, Huang SY, Lai WP, Chou JC. A web-based computerized adaptive testing (CAT) to assess patient perception of Hospitalization. *J Med Internet Res* 2011;13 (3): e61.
- [7] Ma SC, Chien TW, Wang HH, Li YC, Yui MS. Applying computerized adaptive testing to the negative acts questionnaire-revised: Rasch analysis of workplace bullying. *J Med Internet Res* 2014;16 (2): e50.
- [8] Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: University of Chicago Press, 1960.
- [9] Wang WC. Recent Developments in Rasch Measurement. Hong Kong: The Hong Kong Institute of Education Press; 2010.
- [10] Andrich D. A rating scale formulation for ordered response categories. *Psychometrika* 1978;43: 561-573.
- [11] Masters GN. A Rasch model for partial credit scoring. *Psychometrika* 1982;47: 149-174.
- [12] Raîche G, Blais JG, Riopel MA. SAS solution to simulate a Rasch computerized adaptive test. *Rasch Meas Trans* 2006;20 (2): 1061.
- [13] Linacre, JM. Computer-adaptive tests (CAT), standard errors and stopping rules. *Rasch Meas Trans* 2006;20 (2): 1062.
- [14] Chien TW, Wang WC, Lin SB, Lin CY, Guo HR, Su SB. KIDMAP, a web based system for gathering patients' feedback on their doctors. *BMC Med Res Methodol* 2009;9: 38.
- [15] Crossley J, Davies H. Doctors' consultations with children and their parents: a model of competencies, outcomes and confounding influences. *Med Educ* 2005;39 (8): 807-819.
- [16] Davies A, Ware J. Involving consumers in quality of care assessment. *Health Aff (Millwood)* 1988;7 (1): 33-48.
- [17] Levine A. Medical professionalism in the new millennium: a physician charter. 2002;136: 243-226.
- [18] Epstein R, Hundert E. Defining and assessing professional competence. *JAMA* 2002;287: 226-235.
- [19] Maudsley R, Wilson D, Neufeld V, Hennen B, DeVillaeer M, Wakefield J. Educating future physicians for Ontario: phase II. *Acad Med* 2000;75: 113-126.
- [20] Delbanco T. Enriching the doctor-patient relationship by inviting the patient's perspective. *Ann Intern Med* 1992;116: 414-418.
- [21] Hall W, Violato C, Lewkonja R, et al. Assessment of physician performance in Alberta: the Physician Achievement Review. *CMAJ* 1999;161 (1): 52-57.
- [22] Hearnshaw H, Baker R, Cooper A, Eccles M, Soper J. The costs and benefits of asking patients their opinions about general practice. *Fam Pract* 1996;13 (1): 52-58.
- [23] Violato C, Lockyer J, Fidler H. Multisource feedback: a method of assessing surgical practice. *BMJ* 2003;326: 546-548.
- [24] Teutsch C. Patient-doctor communication. *Med Clin North Am* 2003;87 (5): 1115-1145.
- [25] Crossley J, Eiser C, Davies HA. Children and their parents assessing the doctor-patient interaction: a rating system for doctors' communication skills. *Med Educ* 2005;39 (8): 820-828.
- [26] Beckett MK, Elliott MN, Richardson A, Mangione SR. Outpatient satisfaction: the role of nominal versus perceived communication. *Health Serv Res* 2009;44 (5 Pt 1): 1735-1749.
- [27] Linacre JM. WINSTEPS [computer program]. Chicago, IL: <http://www.Winsteps.com>, 2014.
- [28] Smith EV. Detecting and evaluation the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 2002;3: 205-231.
- [29] Linacre JM. User's Guide to Winsteps. Chicago: Mesa Press; 2014.
- [30] Tennant A, Pallant JF. Unidimensionality matters! (A tale of two Smiths?). *Rasch Meas Trans* 2006;20 (1): 1048-1051.
- [31] Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965; 30 (2): 179-185.
- [32] Birnbaum A. Some latent ability models and their use in inferring an examinee's ability. In Lord FM, Novick MR (eds.), *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley; 1968.
- [33] Embretson S, Reise S, Reise SP. *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum; 2000.
- [34] Hsueh IP, Chen JH, Wang CH, Hou WH, Hsieh CL. Development of a computerized adaptive test for assessing activities of daily living in outpatients with stroke. *Phys Ther* 2013;93 (5): 681-693.
- [35] Linacre JM. How to simulate Rasch data. *Rasch Meas Trans* 2007;21 (3): 1125.
- [36] Chien TW. Cronbach's alpha with the dimension coefficient to jointly assess a scale's quality. *Rasch Meas Trans* 2012;26 (3): 1379.
- [37] Chien TW, Wu HM, Wang WC, Castillo RV, Chou W. Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: tool development and simulation. *Health Qual Life Outcomes* 2009: 39.
- [38] Wainer HW, Dorans NJ, Flaugher R, et al. *Computerized adaptive testing: a primer*. Hillsdale, NJ: Erlbaum; 1990.
- [39] Eastaugh SR. Cost containment for the public health. *J Health Care Finance* 2006;32: 20-27.

- [40] Fliege H, Becker J, Walter OB, Bjorner JB, Klapp BF, Rose M. Development of a computer-adaptive test for depression (D-CAT). *Qual Life Res* 2005;14 (10): 2277-2291.
- [41] Chien TW, Wang WC, Huang SY, Lai WP, Chou JC. A web-based computerized adaptive testing (CAT) to assess patient perception of hospitalization. *J Med Internet Res* 2011;13 (3): e61.
- [42] Ma SC, Chien TW, Wang HH, Li YC, Yui MS. Applying computerized adaptive testing to the negative acts questionnaire-revised: Rasch analysis of workplace bullying. *J Med Internet Res* 2014;16 (2): e50.
- [43] Teutsch C. Patient-doctor communication. *Med Clin North Am* 2003;87 (5): 1115-1145.
- [44] Crossley J, Eiser C, Davies HA. Children and their parents assessing the doctor-patient interaction: a rating system for doctors' communication skills. *Med Educ* 2005;39 (8): 820-828.
- [45] Beckett MK, Elliott MN, Richardson A, Mangione SR. Outpatient satisfaction: the role of nominal versus perceived communication. *Health Serv Res* 2009;44 (5 Pt 1): 1735-1749.
- [46] Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas* 2002;3 (1): 85-106.
- [47] Mitchell SJ, Godoy L, Shabazz K, Horn IB. Internet and mobile technology use among urban African American parents: survey study of a clinical population. *J Med Internet Res* 2014;16 (1): e9.
- [48] Linacre JM. RUMM2020 item-trait chi-square and Winsteps DIF size. *Rasch Meas Trans* 2007;21 (1): 1096.