

RESEARCH ARTICLE

Open Access



# Development of a Microsoft Excel tool for one-parameter Rasch model of continuous items: an application to a safety attitude survey

Tsair-Wei Chien<sup>1,2</sup>, Yang Shao<sup>3</sup> and Shu-Chun Kuo<sup>4,5,6\*</sup>

## Abstract

**Background:** Many continuous item responses (CIRs) are encountered in healthcare settings, but no one uses item response theory's (IRT) probabilistic modeling to present graphical presentations for interpreting CIR results. A computer module that is programmed to deal with CIRs is required.

To present a computer module, validate it, and verify its usefulness in dealing with CIR data, and then to apply the model to real healthcare data in order to show how the CIR that can be applied to healthcare settings with an example regarding a safety attitude survey.

**Methods:** Using Microsoft Excel VBA (Visual Basic for Applications), we designed a computer module that minimizes the residuals and calculates model's expected scores according to person responses across items. Rasch models based on a Wright map and on KIDMAP were demonstrated to interpret results of the safety attitude survey.

**Results:** The author-made CIR module yielded OUTFIT mean square (MNSQ) and person measures equivalent to those yielded by professional Rasch Winsteps software. The probabilistic modeling of the CIR module provides messages that are much more valuable to users and show the CIR advantage over classic test theory.

**Conclusions:** Because of advances in computer technology, healthcare users who are familiar to MS Excel can easily apply the study CIR module to deal with continuous variables to benefit comparisons of data with a logistic distribution and model fit statistics.

**Keywords:** Rasch model, Continuous item response, Wright's map, KIDMAP, Visual basic for application

## Background

A great deal of work has been devoted to the probabilistic modeling of Likert-type responses in the past 50 years [1]. Questionnaires built and tested using item response theory (IRT) are common in educational assessment, and they are being used more frequently in health assessment [2]. However, collecting continuous responses is more often and prevalent in healthcare settings than is

collecting categorically ordered data such as questionnaire surveys.

Many researchers have proposed IRT-based models of continuous item responses to deal with those multivariate behavioral research in real world [3–9], no one has provided user-friendly healthcare software for assessing, for example, pain intensity [10] or personality [4, 11]. We were, therefore, interested in developing a Microsoft Excel-based computer module that executes continuous observations similar to easily deal with binary and polytomous scores based on IRT modeling [12, 13].

Samejima [6] developed a unidimensional latent trait model for continuous responses, but it did not aim at “specifically objective” comparisons of persons and

\* Correspondence: rasch.smile@gmail.com

<sup>4</sup>Department of Ophthalmology, Chi-Mei Medical Center, Yong Kang, Tainan City, Taiwan

<sup>5</sup>Department of Optometry, Chung Hwa University of Medical Technology, Jen-Teh, Tainan City, Taiwan

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



items, which is a key and unique feature of Rasch models [14]. Behar [11] applied the normal ogive (logistic curve) case of the continuous response model to items from a personality inventory. Muller [3] proposed a continuous rating scale model, based on Andrich's [15] rating scale model (RSM) for categorical responses, which belongs to the Rasch family and so has the interesting specific objectivity property. Although Muller's model is limited to an "integer" response format, observations can be presented for Rasch analysis in the form of a range of percentages (0–100) without decimal values. Ferrando [4] proposed a similar idea of a truncation mechanism using R language software: extending the linear response model to a nonlinear congeneric model that takes the bounded nature of the data into account rather than rescaling a latent response to fit the response format. This study creates a model in MS Excel that will take the bounded nature of the data into account and rescale a latent response to fit the response format. A unidimensional latent trait model for continuous responses that is easy and friendly use for ordinary practitioners is thus required.

There are two forms for continuous and percentage observations [16]: (i) they are very rarely exist in additive, such as continuous form (e.g., weight and height) and allow us using standard statistical techniques to make direct comparisons between groups; and (ii) they are frequently seen in the real world and might not be directly additive (e.g., time to perform a task and weight lifted with the left hand) or follow a logistic distribution. Their implications in the specific context are unlikely to be additive until they have been transformed to a linear interval score, for instance, using Rasch modeling technique.

IRT-based Rasch probabilistic modeling is a way to transform ordinal scores into interval Rasch measures [2, 3, 9, 12–16], but this type of transformation applies only to discretely ordered category scores. A computer module that is required to deal with continuous item responses (CIRs) that are often seen in healthcare settings. Accordingly, we are interested in rescaling CIRs in a percentage range from 0 to 1 using Rasch probability theory to program a Microsoft Excel-based computer module.

The first section of this article presents the designed response mechanism using the Newton-Raphson method, from which the parameters of items and persons can be estimated and well calibrated. Some resulting estimations yielded by the author-made computer module are verified with the professional Rasch Winsteps software [17] on two discretely ordered category scales [12, 13]. The second applies the model to real healthcare data and shows CIR analysis is more advantageous than is classic test theory (CTT) because the CIR uses probabilistic modeling to deal with continuous and percentage observations.

Based on the study motivation provided, this study aims to accomplish the following:

- i. to verify the CIR that can be a tool used in healthcare settings.
- ii. to apply the CIR to a safety attitude survey using the model's fit statistics.
- iii. to demonstrate an online computer adaptive testing (CAT) of safety attitudes for collecting data from hospital employees.

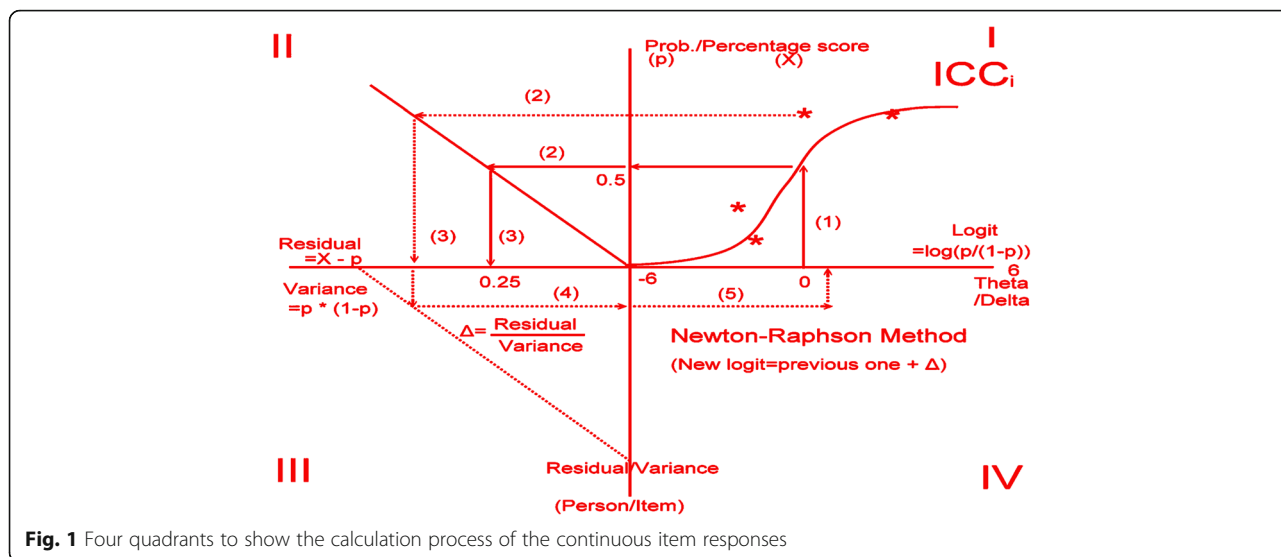
## Methods

### Module design for the continuous item responses (CIRs)

Using Microsoft Excel VBA (Visual Basic for Applications), we designed a computer module that minimizes

the residuals (i.e.,  $\text{Min} \sum_{n=1}^N \sum_{j=1}^L (O_{nj} - E_{nj})$ , where  $N = \text{sample size}$ ,  $L = \text{item length}$ ,  $O = \text{observed score}$ , and  $E = \text{expected value}$ ) for estimating item difficulties parameters ( $\delta_j$ ) and person ability ( $\vartheta_n$ ). The five module steps are illustrated below:

- i) Rescaling the variable scores to a percentage response [i.e., recorded response = (observed response – minimum response)/(maximum response – minimum response)], and setting both initial  $\vartheta_n$  and  $\delta_j = 0$  in a logit unit. This is, at the origin of the measurement scale, every person is estimated to be at the same level of ability, and each item is estimated to be equally difficult.
- ii) Calculating the initial expected values for each response using the one-parameter Rasch probability (Prob) equation:  $\text{Prob} = \exp(\vartheta_n - \delta_j) / (1 + \exp(\vartheta_n - \delta_j)) = \exp(0 - 0) / (1 + \exp(0 - 0)) = 1 / (1 + 1) = 0.5$  (see Quadrant I in Fig. 1).
- iii) Computing variances ( $\text{Var}_{nj} = p_{nj} \times (1 - p_{nj})$ ) and residuals ( $R_{nj} = X_{nj} - p_{nj}$ ) for each response (see Quadrant II in Fig. 1).
- iv) Estimating the provisional  $\theta$  for the row examinees  $\left( = \text{previous } \vartheta + \sum_{j=1}^L R_{nj} / \sum_{j=1}^L \text{Var}_{nj} \right)$  and  $\delta$  for the column items  $\left( = \text{previous } \delta_j - \sum_{n=1}^N R_{nj} / \sum_{n=1}^N \text{Var}_{nj} \right)$  using the Newton-Raphson Method (see Quadrants III and IV in Fig. 1), whereas the mean of  $\delta_j$  is set at zero (i.e.,  $\sum_{j=1}^L \delta_j = 0$ ) [1].
- v) Yielding a newly revised expectation (exp) for each response according the Rasch probability equation [i.e.,  $\text{Prob} = \exp(\vartheta_n - \delta_j) / (1 + \exp(\vartheta_n - \delta_j))$ ] and then doing the loop through the iteration process



from Quadrant II through IV in Fig. 1 until reaching the minimized (Min) sum square residuals (i.e.,  $\text{Min} \sum_{n=1}^N \sum_{j=1}^L (O_{nj} - E_{nj})^2$ ) at a level of 0.05, 0.01, or less.

**Rasch fit statistics used in this study**

In addition to the person measures and item difficulties estimated in our proposed computer module, Rasch mean square (MNSQ) fit statistics (consisting of both item index and person INFIT and OUTFIT indices) describe the fit to the model. The MNSQ fit statistics have a  $\chi^2$  distribution and an expected value of 1, where fit statistics greater than 1 can be interpreted as demonstrating that the observed scores are aberrant distortions of (or underfit) the model's expectation of unidimensionality [18]. Conversely, fit statistics less than 1 can be explained as items (or persons) that overfit the model, or that are highly correlated (i.e., item loading in an explorative factor analysis) with the summation score.

The INFIT MNSQ is referred to as the weighted MNSQ and OUTFIT MNSQ is referred to the unweighted MNSQ. Both are derived from the squared standardized residuals [ $Z_{nj} = \sqrt{R_{nj}^2 / \text{Var}_{nj}}$ , whereas  $\text{Var}_{nj}$  are response variances, and the person (n) and item (j) OUTFIT MNSQs are presented as  $\sum_{j=1}^L Z_{nj}^2 / L$

and  $\sum_{n=1}^N Z_{nj}^2 / N$ ] for each item/person interaction. The

OUTFIT MNSQ is the average of the standardized residual variance across items and persons; it is unweighted, which means that the estimate produced is relatively more affected by unexpected responses distant to item or person measures [19]. For the INFIT MNSQ

the residuals are weighted by their individual variance

$$\text{[i.e., } \frac{\sum_{n=1}^N (R_{nj})^2}{\sum_{n=1}^N \text{Var}_{nj}} \text{ for the item (j) and } \frac{\sum_{j=1}^L (R_{nj})^2}{\sum_{j=1}^L \text{Var}_{nj}} \text{ for the}$$

person (n)] to minimize the effect of unexpected responses far from the measure.

The INFIT MNSQ is relatively more affected by unexpected responses closer to item and person measures [18]. In this study, the INFIT MNSQs for items less than 1.5 means data fitting to the Rasch model of the unidimensionality requirement and the OUTFIT MNSQs for persons less than 2.0 indicates person responses without aberrant or careless behaviors [20]. Additionally, the squared standardized residuals ( $Z_{nj} = \sqrt{R_{nj}^2 / \text{Var}_{nj}}$ ) for each response less than 2.0 (i.e.,  $p < 0.05$ ) is defined as a nonsignificant deviation from the model's expected value.

**Validation of the CIR model**

Besides the aforementioned theoretical modelling method, two ordered-category datasets derived using the 11-item dichotomous and 26-item polytomous scales [12, 13] were further used to verify whether our proposed computer module yield accurate results in measures, difficulties, and fit statistics similar to those of the professional Rasch Winsteps software [17]. The percentage scores were rescaled to a range between 0 and 1.0. For instance, the only two values of 0 and 1.0 were assigned the original codes of 0 and 1, and the 5-point Likert-type scaling scores were uniformly transformed to 0, 0.25, 0.5, 0.75, and 1.0, respectively.

The 2-item variables downloaded from a 2015 US NBA game between GSW (Golden State Warriors) and OKC (Oklahoma City Thunder) on Feb. 27, 2016 were used for showing the CIR that transforms raw scores into Rasch estimated logits.

### Demonstrations of real healthcare data using the Rasch CIR model

Two published papers [21, 22] were referenced to evaluate whether our proposed Rasch CIR model using the fit statistics and the expected responses is more beneficial than the traditional CTT method because it discloses more valuable messages. One data source was collected from a Taiwan hospital in a safety attitude survey [22, 23]. An online computer adaptive testing (CAT) that can deal with many CAT procedural on each domain in a test was demonstrated to collect safety attitude data from hospital employees.

### Statistical analysis

In the author-made CIR module,  $R^2$  was calculated using the regression equation  $\left( = \sum_{n=1}^N \sum_{j=1}^L (E_{nj} - \bar{X}_{ij})^2 / \sum_{n=1}^N \sum_{j=1}^L (X_{nj} - \bar{X}_{ij})^2 \right)$ . The Paired- $t$  and independent- $t$  tests for two variables were used to compare the raw scores, percentage scores, and Rasch expected response scores using the 2-variable dataset of safety attitudes. The person OUTFIT MNSQ was obtained using the two values of standardized residual ( $Z$ ) and item length ( $L$ ) in the equation  $\left( = \sum_{j=1}^L Z_{nj}^2 / L \right)$ . The Wright Map (or named person-item map) [24] and the KIDMAP [25] were drawn using the CIR module.

## Results

### Validation of the Rasch CIR model

The Outfit MNSQ statistics and measures yielded by the proposed CIR computer module and by the Rasch professional Winsteps software [17] on the two study scales [12, 13] were equivalent (Fig. 2). We can see that spot lines of the CIR and Winsteps software in 4 combinations of Outfit MNSQ and measures on two scales (i.e., dichotomous and polytomous responses) are closely overlaid [26].

The CIR was applied to the 2-item competition basketball scores between GSW and OKC on Feb. 27, 2016. Two variables' item character curves (ICC) with person Outfit MNSQ are present in Fig. 3 (left panel) after transforming raw scores into linear logits. It can be evident of the fact that logit scores (on X axle) are more appropriate than those raw scores (on Y axle) when adding them for group comparisons and release more information, such as person Outfit MNSQ, to readers on a

plot. The more difficult items (i.e., lower scores) are shown to the right-hand side. The more aberrant responses (i.e., response pattern not consistent to others) are shown to the top with Outfit MNSQ greater than 2.0 [20] on a plot (left panel in Fig. 3).

### Real data used to release information with visual presentations

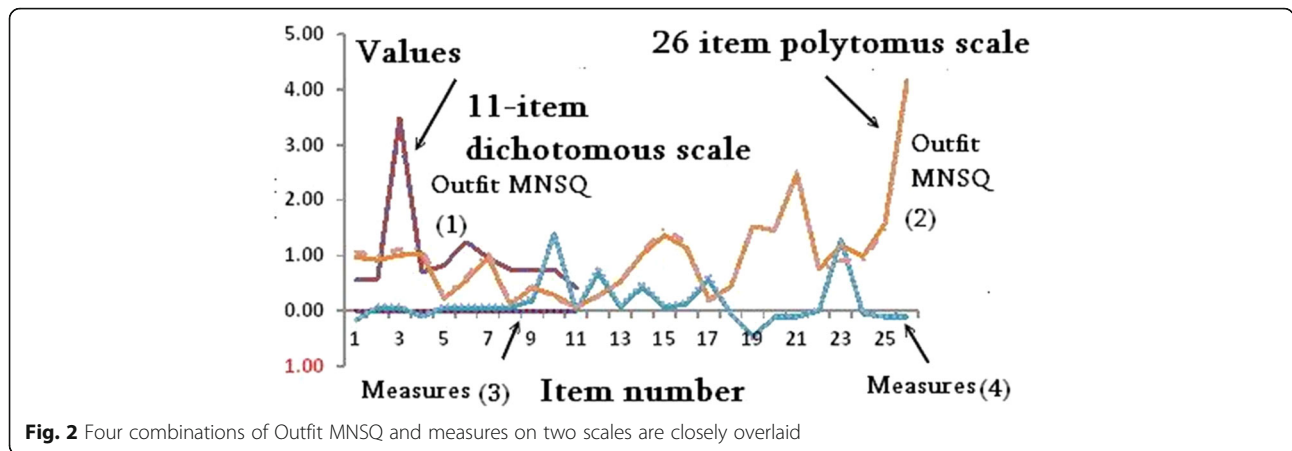
Besides the theoretical modelling method verified in Fig. 1, the accurate measures estimated proved in Fig. 2, and the logit scores transformed by the CIR in Fig. 3 (on X axle), we used the survey resulting data from a Taiwan medical center with 1200 beds in 2015 to demonstrate the application of the CIR  $t$  (see Additional file 1 in Appendix). A total of 80 departments were included to show the 6 domain scores (including Teamwork Climate, Safety Climate, Job Satisfaction, Stress Recognition, Perception of Management, and Working Condition) [21, 22, 27, 28] on a plot (right panel in Fig. 3). All domain scores are above a percentage of 0.3 (equivalent to 30 points in a 0–100 scale), two domains (i.e., Perception of Management and Working Condition, not present a linear pattern with a R-square around 0.98) are more difficult than other four, and several aberrantly unexpected responses are present in Figure (right) with Outfit MNSQ greater than 2.0 [20].

The data metric of 80 units \* 6 domains were analyzed with the model-data fit (see Additional file 2 in Appendix).  $R^2$  was 0.65 with  $F = 1.96$  ( $p = 0.18$ ) using the regression equation  $\left( = \sum_{n=1}^N \sum_{j=1}^L (E_{nj} - \bar{X}_{ij})^2 / \sum_{n=1}^N \sum_{j=1}^L (X_{nj} - \bar{X}_{ij})^2 \right)$ . The  $t$ -value of the two item difficulties ( $-0.23$  and  $-0.42$  logits for Teamwork Climate and Safety Climate) was  $0.48$   $\left( = (-0.23 - -0.42) / \sqrt{1/13.76 + 1/12.39} \right)$ ,  $p = 0.63$ , not significantly different from each other using the traditional  $t$ -test.

A Wright map shows that the 6 domains and a single construct when referring to INFIT MNSQs  $< 1.5$ ; the person distribution is also shown (Fig. 4, left).

The results of a case department show an aberrant OUTFIT MNSQ of 5.67 ( $> 2.0$ ) in the Box KIDMAP (Fig. 5, right bottom). One domains—Safety Climate—shows a significantly high Z-score ( $= 2.75$ ):  $> 2.0$ , and another dimension—Stress Recognition—shows a significantly low Z-score ( $= -4.55$ ):  $< -2.0$ , implying the strength and weak comparing to other counterparts.

From the Box KIDMAP we also see that the most difficult dimension (i.e., the one with the lowest mean score) was Perception of Management and Working Condition. The easiest one (the one with the highest mean scores) was Safety Climate. Most of the dimension



scores appear to be normally distributed, but Job Satisfaction was slightly negatively distributed.

The most unexpected misfit case—an OUTFIT MNSQ 5.56 (>2.0)—was extracted at the 1.28 logits point (Fig. 6). We can see the aberrant response pattern with a high OUTFIT MNSQ (>2.0) (e.g., this anomaly might be most attributable to the two domains with the two Z-scores of 2.75 and -4.55 in Fig. 5).

**An online computer adaptive testing (CAT) was demonstrated**

By scanning the QR code (Fig. 7, middle right), the CAT icon appears on the patient’s mobile phone. We demonstrate an alternative way for gathering respondent’s safety attitudes with item-by-item in action (see Additional file 3 in Appendix). Person fit (ie, infit and outfit mean square) statistics showed the respondent behaviors. Person theta is the provisional ability estimated by the online CAT module.

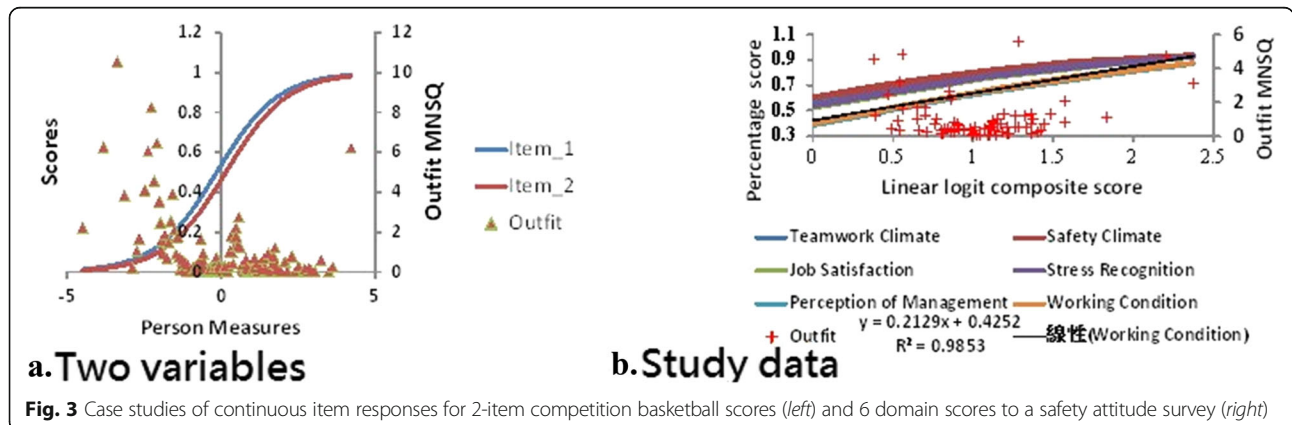
The MSE (e.g., measurement of standard error) in Fig. 7 was generated by the following formula [29]:  $1/\sqrt{(\sum \text{variance}(i))}$ , where  $i$  refers to the CAT finished items responded to by a person [30]. The residual (resi) on top in Fig. 7 was

the average of the last three change differences between the pre-and post-estimated abilities on each CAT step. CAT will stop if the residual value is less than 0.05, indicating a stable estimation in process. Corr refers to the correlation coefficient between the CAT estimated measures and its step series numbers using the last three estimated theta (= person measure) values. The flatter the theta trend, the higher the probability that the person measure is convergent with a final estimation.

**Discussion**

**Principal findings**

Our most important finding was that the OUTFIT MNSQ and person measures yielded by our proposed CIR module were not significantly different from the professional Rasch Winsteps software on the two dichotomous and polytomous scales. Our second most important finding was that the probabilistic modeling of our CIR module presented readers much more valuable messages (e.g., fit statistics and visual representations of a Wright map and KIDMAP) and illustrated the CIR module’s advantage over CTT, especially when using the expected scores (instead of the raw scores) in comparing group (or item) differences.



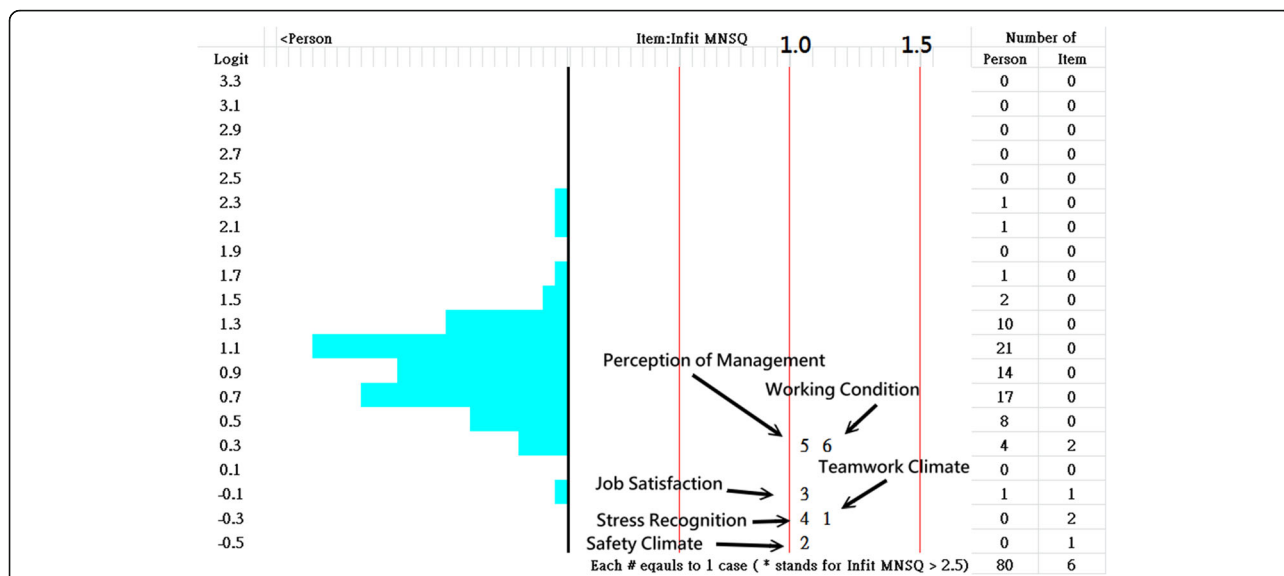


Fig. 4 Wright map shows person and item dispersions

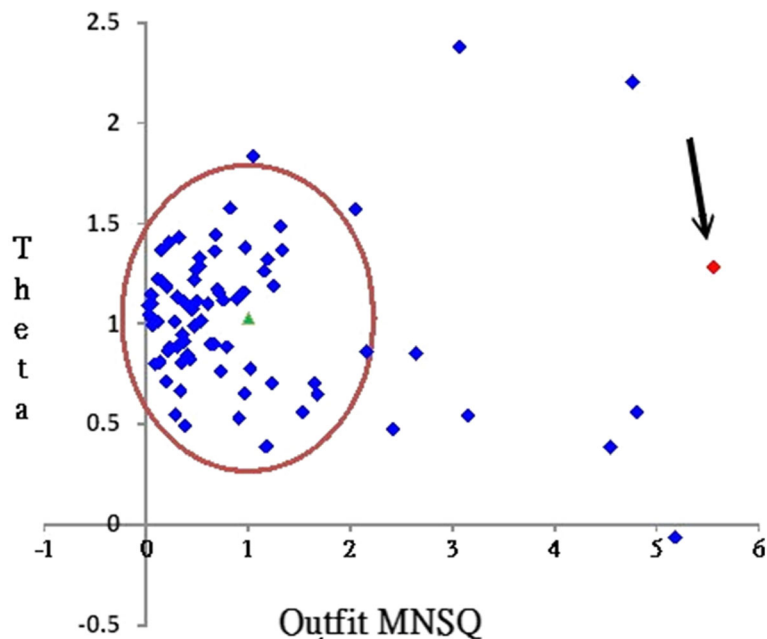
A video regarding the CIR application is demonstrated in a MP4 format at reference [31].

There are so many comparative scores that are not, or might not be, additive but that follow a logistic distribution in the real world [16]. The Rasch model formulates a nonlinear relationship between nonlinear raw scores and additive measures [31]. Figure 4 shows that we were

more concerned about locating that person on the true score trait (on X axle) than in having that person obtains a raw score (on Y axle). This is why we preferred to use the model-expected scores for comparison in Fig. 1. Unfortunately, we have not seen a statistical modeling technique that can deal with CIRs as we did using our Microsoft Excel module.



Fig. 5 Box-plot KIDMAP shows a person performance on a sheet



**Fig. 6** A person performance referred to the overall counterparts

### Implications and future considerations

We showed that fit statistics and visual representations of a Wright map and KIDMAP offer much more valuable messages to readers. The premises consist of the model-expected scores that should first be yielded by Rasch analysis. Fit statistics are thus produced, and the Wright map and KIDMAP follow. An adequate CIR computer routine is thus required.

Using our proposed module in the future will not be limited to the two types of continuous item responses (i.e., two basketball completion scores and six items of the patient safety survey shown in Fig. 4) illustrated here. Scores for a single item (e.g., serum levels of liver enzymes and lipids [32]) can also be easily transformed into Rasch interval logit measures used for statistical group comparison or product moment correlation analysis in practice.

Our CIR computer program can be simultaneously used for many between-test dimension scores in a survey, like in this study. For example, the Clinical Dementia Rating scale [33] used for patients with dementia consists of six domains. We can present the Wright map and KIDMAP to the patient or their family proxy in the same way that educational performance sheets are presented to students or their parents [24, 25].

### Strengths

All Rasch measurement analysis software is used for dichotomous and polytomous ordered category scales [33, 34]. In the healthcare field, there are so many continuous variables that cannot be analyzed and plotted a Wright map and KIDMAP. Many studies [3–9] have proposed IRT-based

models used for CIRs but have not provided user-friendly software such as MS Excel version in this study. We demonstrated one example of a safety attitude survey using continuous variables analyzed with a user-friendly MS Excel module to plot a Rasch Wright map and KIDMAP.

The Wright map was improved by adding the item INFIT MNSQ dispersion on the right (see Fig. 4) and with color features, which has never before been used to interpret the unidimensionality feature together with person distribution in a diagram.

The KIDMAP has evolved to a colorful item-by-item bar-chart plot (see Fig. 5) different from the traditional monochrome diagram [35, 36–38] or KIDMAP solely for dichotomous scales [25]. More information has been included in the colorful KIDMAP: all person distributions and dispersions along all items, Z-scores displaying the most unexpected items, and the individual person location points on items suitable for polytomous and continuous variables. A person performance referred to the overall counterparts is also present in a scatter plot (Fig. 6) that can easily select aberrant respondents for further improvement and concerns with respect to the example of hospital safety attitude survey.

It is worth mentioning that the plot combined model-expected item character curves with OUTFIT MNSQs in a diagram (Fig. 3) which is featured and outlined together to present many valuable messages that are not easily disclosed using the traditional CTT method. The between-test score (i.e., CIR format) comparison demonstrated in this study is an extension of those within-test score analyses (i.e., ordered category format). For instance, we concern with outcomes

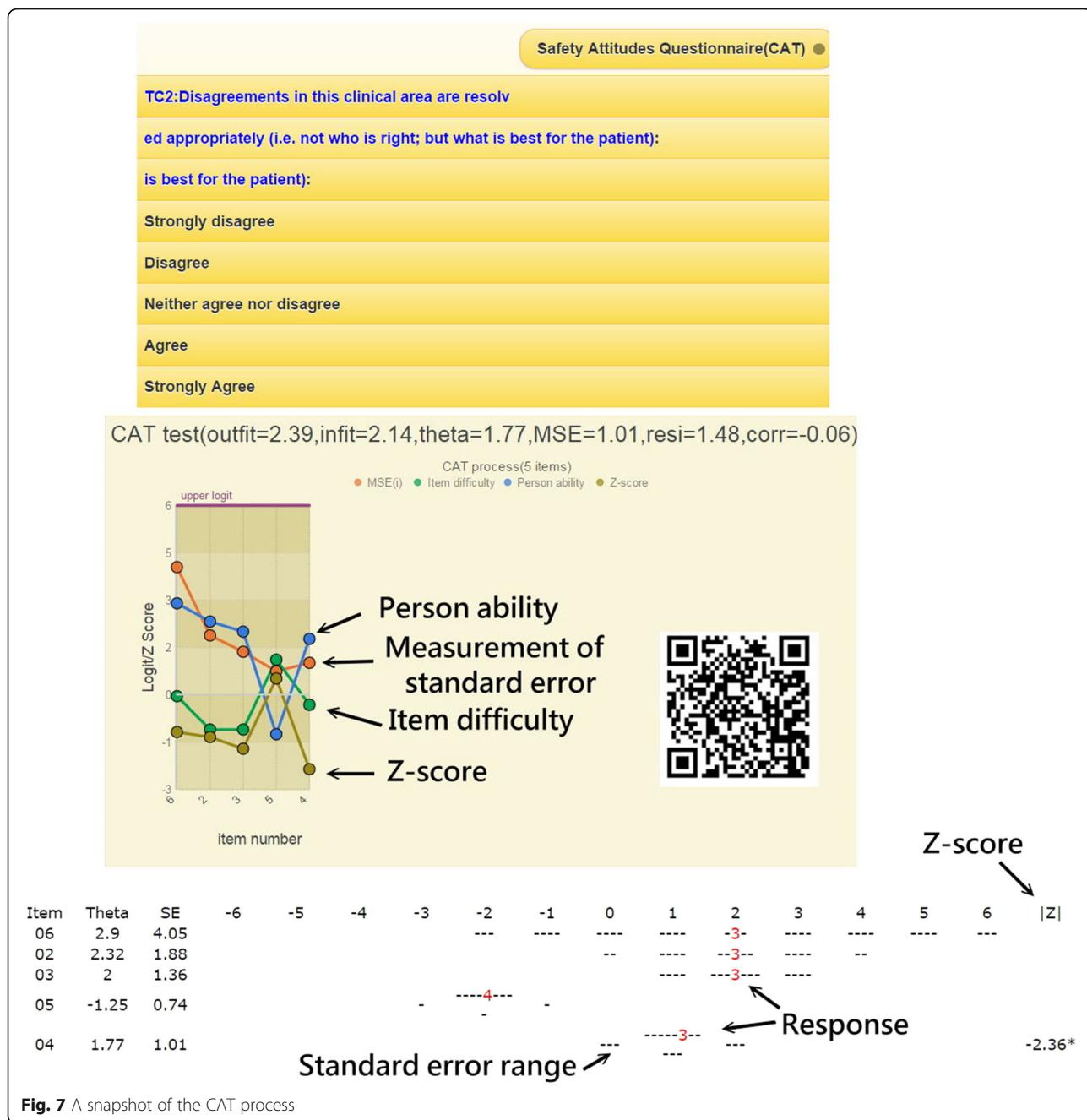


Fig. 7 A snapshot of the CAT process

between groups (with continuous format) after gathering data on the individual base (with ordered category format).

Many studies [39–43] have reported that item response theory (IRT)-based computer adaptive testing (CAT) has the advantages of both long-form and short-form questionnaires [44–46] in precision and efficiency. Many employees already own mobile phones. No studies reported online CAT via mobile phones in medical fields, especially implementing multiply-domain online CAT together in a test. We demonstrated it for gathering survey data regarding safety attitudes here (See Fig. 7).

**Limitations**

This study has at least four limitations. First, the residuals in person-item interaction cells are significantly small in the CIR format scenario. A normalization procedure for letting the mean of INFIT and OUTFIT MNSQ = 1.0 is necessary for easy interpretation by readers, assuming that the acceptable criterion is between 0.5 and 1.5 required by the Rasch model.

Second, IRT parameters in the online CAT example is retrieved from a published paper [21] using the IRT graded response model (GRM) [47], which did not comply with



“specifically objective” comparisons of persons and items as the unique feature of Rasch models [14]. We can replace all the discrimination parameters in GRM with 1 to fit Rasch model's requirement in future study.

Third, we combined the person OUTFIT MNSQ into the model-expected item character curves in a diagram (Fig. 3) because two vertical axes with a significant difference distort the diagram and make it difficult to interpret without training, but easy to compare the two vertical axes after training.

Fourth, the CIR module is easy to yield greater Outfit MNSQ in the extreme two sides (i.e., toward zero and 100 percentage score) because of minimal variance leading a higher value of person Outfit

MNSQ  $\left( = \sum_{j=1}^L R_{nj}^2 / \text{Var}_{nj} / L = \sum_{j=1}^L Z_{nj}^2 / L \right)$ . For instance, a variance (=0.5\*0.5) in middle point is greater than it (=0.1\*.09) in both extreme sides.

Fifth, we merely emphasize the usefulness on the healthcare settings, not suggest the CIR being applied to the clinical utility. Because the chemical Lab examination scores are not like the survey items having a consistent scaling score. Whether the clinical scores should be standardized first prior to analysis is required to further explored and researched in future.

## Conclusion

Because of advances in computer technology, healthcare users who are familiar to MS Excel can easily apply the study CIR module to deal with continuous variables to benefit comparisons of data with a logistic distribution and model fit statistics.

## Additional files

**Additional file 1:** Microsoft Excel-based computer module for continuous item responses. (XLSM 2705 kb)

**Additional file 2:** Video demonstration regarding application of the Excel CIR module [http://www.healthup.org.tw/marketing/course/information/rasch\\_cir.mp4](http://www.healthup.org.tw/marketing/course/information/rasch_cir.mp4). (MP4 45229 kb)

**Additional file 3:** Computerized adaptive testing of the patient safety survey on six subscales <http://www.healthup.org.tw/marketing/course/information/safetysurveyKREA.mp4>. (MP4 14747 kb)

## Abbreviations

CAT: Computer adaptive testing; CIR: Continuous item responses; CTT: Classic test theory; GRM: Graded response model; ICC: Item character curves; IRT: Item response theory; MNSQ: Mean square error; MSE: Measurement of standard error; RSM: Rating scale model; VBA: Visual basic for applications

## Acknowledgements

We thank Frank Bill who provided medical writing services to the manuscript.

## Funding

There are no sources of funding to be declared.

## Availability of data and materials

All data used for verifying the proposed computer module during this study are extracted from the published articles [12, 13]. The Microsoft Excel-based computer module including the demonstrated data of safety attitudes can be downloaded from the supplementary information files.

## Authors' contributions

TW developed the study concept and design. TW and YS analyzed and interpreted the data. TW drafted the manuscript, and all authors provided critical revisions for important intellectual content. All authors have read and approved the final manuscript as well as agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was supervised by SC.

## Competing interests

The authors declare that they have no competing interests.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

The secondary data were retrieved from published papers on references from 21 to 23, one for CAT used as item pool and another for demonstration in MS Excel. With permission from the CEO of the Joint Commission of Taiwan who is the correspondence author on reference 23, the data of safety attitudes demonstrated were obtained from the published paper, in which all hospital and study participant identifiers were stripped and merely grouped into composite scores of 80 anonymous divisions.

## Author details

<sup>1</sup>Medical Research Department, Chi-Mei Medical Center, Tainan, Taiwan.

<sup>2</sup>Department of Hospital and Health Care Administration, Chia-Nan University of Pharmacy and Science, Tainan, Taiwan. <sup>3</sup>Department of Electronics and Information Engineering, Tongji Zhejiang College, Jiaxing, China.

<sup>4</sup>Department of Ophthalmology, Chi-Mei Medical Center, Yong Kang, Tainan City, Taiwan. <sup>5</sup>Department of Optometry, Chung Hwa University of Medical Technology, Jen-Teh, Tainan City, Taiwan. <sup>6</sup>Chi-Mei Medical Center, No. 901, Chung Hwa Road, Yung Kung Dist, Tainan 710, Taiwan.

Received: 20 October 2016 Accepted: 8 December 2016

Published online: 10 January 2017

## References

- Noel Y, Dauvier B. A Beta item response model for continuous bounded responses. *Appl Psychol Meas*. 2007;31:47.
- Djaja N, Janda M, Olsen CM, Whiteman DC, Chien TW. Estimating Skin Cancer Risk: Evaluating Mobile Computer-Adaptive Testing. *J Med Internet Res*. 2016;18(1):e22.
- Müller H. A Rasch model for continuous ratings. *Psychometrika*. 1987;52(2):165–81.
- Ferrando PJ. A nonlinear congeneric model for continuous item responses. *Br J Math Stat Psychol*. 2001;54:293–313.
- Mellenbergh GJ. A unidimensional latent trait model for continuous item responses. *Multivar Behav Res*. 1994;29:223–36.
- Samejima F. Homogeneous case of the continuous response model. *Psychometrika*. 1973;38:203–19.
- Samejima F. Normal ogive model on the continuous response level in the multidimensional latent space. *Psychometrika*. 1974;39:111–21.
- Wang T, Zeng L. Item parameter estimation for a continuous response model using an EM algorithm. *Appl Psychol Meas*. 1998;22:333–44.
- Linacre JM. Percentages with continuous Rasch models. *Rasch Meas Trans*. 2001;14(4):771–4.
- Morin C, Bushnell MC. Temporal and qualitative properties of cold pain and heat pain: A psychophysical study. *Pain*. 1998;4:67–73.
- Bejar II. An application of the continuous response model to personality measurement. *Appl Psychol Meas*. 1977;1:509–21.
- Linacre JM. (Dichotomous Mean-square) Chi-square fit statistics. *Rasch Meas Trans*. 1994;8(2):360.

13. Linacre JM. An All-Purpose Person Fit Statistic? *Rasch Meas Trans.* 1997;11(3):582–3.
14. Rasch G. Probabilistic models for some Intelligence and attainment tests. Chicago: University of Chicago Press; 1960.
15. Andrich D. A rating formulation for ordered response categories. *Psychometrika.* 1978;43:561–73.
16. Linacre JM. Decimal, percentage and continuous data. retrieved at <http://www.winsteps.com/winman/decimal.htm>. Accessed 20 Feb 2016.
17. Linacre JM. Winsteps software. retrieved at <http://www.winsteps.com>. Accessed 20 Feb 2016.
18. Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Hillsdale: Lawrence Erlbaum Baum Associates; 2001.
19. Smith A, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Med Res Methodol.* 2008;8:33.
20. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas.* 2002;3(1):85–106.
21. Jeong HJ, Lee WC. Item response theory-based evaluation of psychometric properties of the safety attitudes questionnaire—Korean version (SAQ-K). *Biometrics Biom Biostat Int J.* 2016;3(5):00079.
22. Lee WC, Wung HY, Liao HH, Lo CM, Chang FL, Wang PC, Fan A, Chen HH, Yang HC, Hou SM. Hospital safety culture in Taiwan: a nationwide survey using Chinese version Safety Attitude Questionnaire. *BMC Health Serv Res.* 2010;10:234.
23. Chien TW, Chang Y, Chien PS, Lin HJ. A dashboard used for displaying the results of the hospital patient safety culture survey. *J Taiwan Assoc Med Inform.* 2015;24(4):49–59.
24. Wilson M. Some Notes on the Term: "Wright Map". *Rasch Meas Trans.* 2011; 25(3):1331.
25. Masters GN. KIDMAP – a history. *Rasch Meas Trans.* 1994;8(2):366.
26. Chien TW, Shao Y. Rasch analysis for continuous variables. *Rasch Meas Trans.* 2016;30(1):1574–6.
27. Jeong HJ, Jung SM, An EA, Kim SY, Yoon H. Development of the safety attitudes questionnaire-Korean version (SAQ-K) and Its novel analysis methods for safety managers. *Biometrics Biostatistics Int J.* 2015;12(1):1–20.
28. Jeong HJ, Jung SM, An EA, Kim SY, Song BJ. A strategy to develop tailored patient safety culture Improvement Programs with Latent Class Analysis Method. *Biometrics Biostatistics Int J.* 2015;2(2):1–27.
29. Chien TW, Lin WS. Improving Inpatient Surveys: Web-Based Computer Adaptive Testing Accessed via Mobile Phone QR Codes. *JMIR Med Inform.* 2016;4(1):e8.
30. Linacre JM. Computer-adaptive tests (CAT), standard errors and stopping rules. *Rasch Meas Trans.* 2006;20(2):1062.
31. Linacre JM. Iterations - PROX & JMLE. Winsteps® Rasch measurement computer program User's Guide. Beaverton: Winsteps.com; 2016.
32. Alcalay J, Landau M, Zucker A. Analysis of laboratory data in acne patients treated with isotretinoin: is there really a need to perform routine laboratory tests? *J Dermatolog Treat.* 2001;12(1):9–12.
33. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *Br J Psychiatry.* 1982;140(6):566–72.
34. Wang WC, Chen CT. Item parameter recovery, standard error estimates, and fit statistics of the Winsteps program for family of Rasch models. *Educ Psychol Meas.* 2005;65(3):376–404.
35. Linacre JM. Rasch measurement analysis software directory. retrieved at <http://www.rasch.org/software.htm>. Accessed 20 Feb 2016.
36. Chien TW, Wang WC, Wang HY, Lin HJ. Online assessment of patients' views on hospital performances using Rasch model's KIDMAP diagram. *BMC Health Serv Res.* 2009;9:135.
37. Chien TW, Wang WC, Roberto VC, Su SB. A graphical health report constructed as a KIDMAP using Rasch analysis of IRT model. *Health Med.* 2012;6(1):29–36.
38. 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.
39. 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;7:39.
40. 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.
41. 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.
42. 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.
43. Stochl J, Böhnke JR, Pickett KE, Croudace TJ. An evaluation of computerized adaptive testing for general psychological distress: combining GHQ-12 and Affectometer-2 in an item bank for public mental health research. *BMC Med Res Methodol.* 2016;16:58.
44. 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):465–74.
45. Ramirez Basco M, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W, et al. Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry.* 2000;157(10):1599–605.
46. Shear MK, Greeno C, Kang J, Ludewig D, Frank E, Swartz HA, et al. Diagnosis of nonpsychotic patients in community clinics. *Am J Psychiatry.* 2000;157(4):581–7.
47. Samejima F. Estimation of latent ability using a response pattern of graded scores. *Research Bulletin, Princeton: Educational Testing Service; 1962.* p. 2.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

