

Research Article

Volume 8 Issue 5 - April 2018
DOI: 10.19080/PBSIJ.2018.08.555750

Psychol Behav Sci Int J

Copyright © All rights are reserved by Chien-ho Lin

Online Computerized Adaptive Testing to Examine a Person Being Depressed Using a Taiwanese Depression Scale (TDS)



Tsair-Wei Chien^{1,2}, Yu Chang³ and Chien-ho Lin^{4*}

¹Research Departments, Chi-Mei Medical Center, Taiwan

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

³National Taiwan University School of Medicine, Taiwan

⁴Department of psychiatry, Chi-Mei Medical Center, Taiwan

Submission: March 10, 2018; **Published:** April 13, 2018

***Corresponding author:** Chien-ho Lin, Chi-Mei Medical Center, No. 901, Chung Hwa Road, Yung Kung Dist, Tainan 710, Taiwan, Email: rasch.smile@gmail.com

Abstract

Background: Person's depression has been measured in many studies to investigate mental health issues. None uses online computerized adaptive testing (CAT) with cutting points to report a prevalence rate of depression at a workplace.

Objective: To develop an online CAT to examine person being depressed and verify whether item response theory-based computerized adaptive testing (CAT) can be online applied to measure person's depression.

Methods: A total of 413 persons (213 depression patients and 200 normal undergraduates) were recruited and responded to the 22-item Taiwanese Depression Scale (TDS). All non-adaptive testing (NAT) items were calibrated with the Rasch rating scale model. Three scenarios (i.e., NAT, CAT, and the randomly-selected method to NAT) were manipulated to compare their response efficiency and precision by comparing i) item length for answering questions, person measure, ii) correlation coefficients, iii) paired t tests, and iv) estimated standard errors (SE) between CAT and the random to its counterpart of NAT.

Results: The TDS is a unidimensional construct that can be applied for patients to measure depressive perceptions on CAT. CAT required fewer items (= 14) than NAT (= 22, an efficient gain of 36% = 1-14/22). Person measures derived from both tests (CAT and the random to NAT) were highly correlated ($r = 0.96$ and 0.98) and their measurement precisions were not statistically different (the percentage of significant count number less than 5%) as expected, but CAT earns substantially smaller person measure SE than the random scenario. The positive and negative predictive values for this study were 0.96 and 0.91, respectively, when cutting points were set at -0.7 and 0.7 logits.

Conclusion: With CAT online-based administration of the TDS for patients, their burden was substantially reduced without compromising measurement precision.

Keywords: Computerized Adaptive Testing; Non-Adaptive Testing; Item Response Theory; Taiwanese Depression Scale; Item Response Theory

Abbreviations: CAT: Computer Adaptive Testing; ICC: Item Character Curves; IRT: Item Response Theory; MNSQ: Mean Square Error; MSE: Measurement of Standard Error; NAT: Non-Adaptive Testing; NIMH: The US National Institute of Mental Health; RSM: Rating Scale Model; SE: Standard Error; SD: Standard Deviation; TDS: Taiwanese Depressive Scale; VBA: Visual Basic for Applications

Introduction

Depression is a disease of modernity with an increasing prevalence rate due to drastic changes in daily life over the past century [1]. Mental disorders are common in the United States and internationally. According to a report from the US National Institute of Mental Health (NIMH) in 2013, an estimated 26.2 percent of Americans ages 18 and older, about one in four adults, suffer from a diagnosable mental disorder in a given year [2],

increasing from 16.2% in 2003 [3]. Major Depressive Disorder (MDD) is the leading cause of disability in the U.S. for ages 15-44 [4]. MDD affects about 14.8 million American adults (i.e., 6.7% of the U.S. population age beyond 18) in a given year [5,6]. While MDD develops at any age, the median age at onset is 32 years old [7]. MDD is more prevalent (about two times) in women than in men [8]. People are most likely to suffer their first depressive

episode between the ages from 30 to 40, and there is a second, smaller peak of incidence between ages 50 and 60 [9].

DSM-III criteria were used to define depression. Lifetime prevalence estimates of MDD ranged from 1.0% (Czech Republic) to 16.9% (US), with midpoints at 8.3% (Canada) and 9.0% (Chile)[10], 1.14% in 1996 [11] and 1.20% in 2012 [12] (Taiwan). Weissman et al. [13] published the first cross-national comparison of major depression from 10 population-based surveys in 1996 and reported their prevalence rates from the lowest Taiwan (1.5%) to highest Beirut (19.0%), with the midpoints at 9.2% (West Germany) and 9.6% (Edmonton, Canada). How to develop an easy and friendly assessment that can help an institute or a school to consecutively monitor its own MDD prevalence rate is urgently required in this high-tech industrial society.

An Assessment to calculate depression prevalence rates

Many kinds of depression scale have been developed and validated in published papers [14-18]. Some were translated into local languages for use and some were developed by researchers using their own languages. All of them encountered a common problem that is to determine cutting points for calculating a depression prevalence rate. Different item length and category number lead to different cutting points if summation score is applied. Importantly, a comparison between derived score levels and the suggested best cutoff points can help clinicians (or practitioners) assess examinees at risk of an incidence [19,20]. Multiple cutoff points are recommended being more powerful and useful than one single cutoff point [21,22]. How to determine appropriate cutting points for a depression scale is the first research question of the current study.

Online computerized adaptive testing is required

Many studies [23-27] have addressed that item response theory (IRT)-based computer adaptive testing (CAT) has the advantages of both long-form and short-form questionnaires

Cutting points used for the TDS

[28-30] in precision and efficiency. Mobile phones are commonly used by people in all walks of life in this technology era. However, no studies till now report any kind of online CAT via mobile phones for gathering data in healthcare settings. The second research question is how to develop an online CAT that can be used for a depression assessment.

Objectives

First, we demonstrate a Taiwanese Depression Scale (TDS) that is a unidimensional construct. Second, we determine a set of cutting points that can be used for computing a prevalence rate at workplace on CAT. Third, we compare CAT with non-adaptive testing (NAT) and the randomly selected method to NAT on efficiency and precision. Fourth, we develop an online CAT for individual examinees to measure the level of person depression.

Methods

Taiwanese depression scale (TDS) and data source

The study data (i.e., item and person parameters) was extracted from two published papers [31,32], comprising 213 depression patients and 200 normal undergraduates who answered 22-item TDS in 4-point Likert-type format (i.e., 0 for seldom, 1 for occasionally, 2 for frequently, and 3 for always). Four facets included in this TDS scale are cognitive dimension with six items; emotional dimension with six items; physical dimension with six items, and interpersonal dimension with four items. It was evident that the 22-item TDS (Cronbach's $\alpha=0.97$) can be a unidimensional instrument for evaluating depressive symptoms and act as a means to replace some out-of-date depression scales in Taiwan [31]. With 22 item difficulties, three threshold difficulties under Rasch [33] rating scale model [34], and 413 person ability parameters, we conducted a simulation [35] to form a 413×22 response rectangle metric fitting to Rasch model's requirement (see the demonstration in Additional files from 1 to 3). Rasch person separation reliability is 0.85 (Mean=-0.70, SD=1.81).

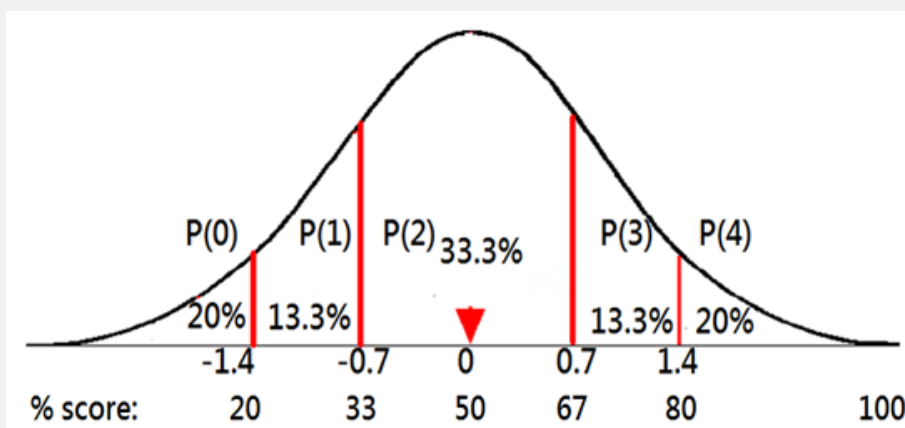


Figure 1: Cutting points at the probability density of a continuous latent random variable.

According to the literature [36-38], as a scale's reliability (i.e., Cronbach's α) increases, so does the person-number of ranges that can be confidently distinguished. Measures from two instruments with reliabilities of 0.67 will tend to vary within two groups that can be separated with 95% confidence; 0.80 will vary within three groups; 0.90, within four groups; 0.94, within five groups; 0.96, within six groups; 0.97, within seven groups etc. [39]. More conservative to compute the number of the strata, the scale reliability was relied on the Rasch person separation reliability ($=0.85$), and then referred to the Rasch threshold

difficulty guideline [40] with an appropriate distance between two thresholds ranging from 1.4 to 5.0 logits (log odds). Three strata were thus determined. Standard error of measurement was $0.7(=SD - reliability=1.81 - 0.85)$. Accordingly, the cutting points can be set at -0.7 and 0.7 logits when the mean of the 22 item difficulties is usually calibrated at zero logit (Figure 1). A comparison was made to select the highest Kappa coefficient and hit ratio (i.e., accurate classification rate=the number of accurate classification in both positive negative and cells divided by the total number) among all possible cutting points.

Comparison of efficiency and precision using CAT algorithm

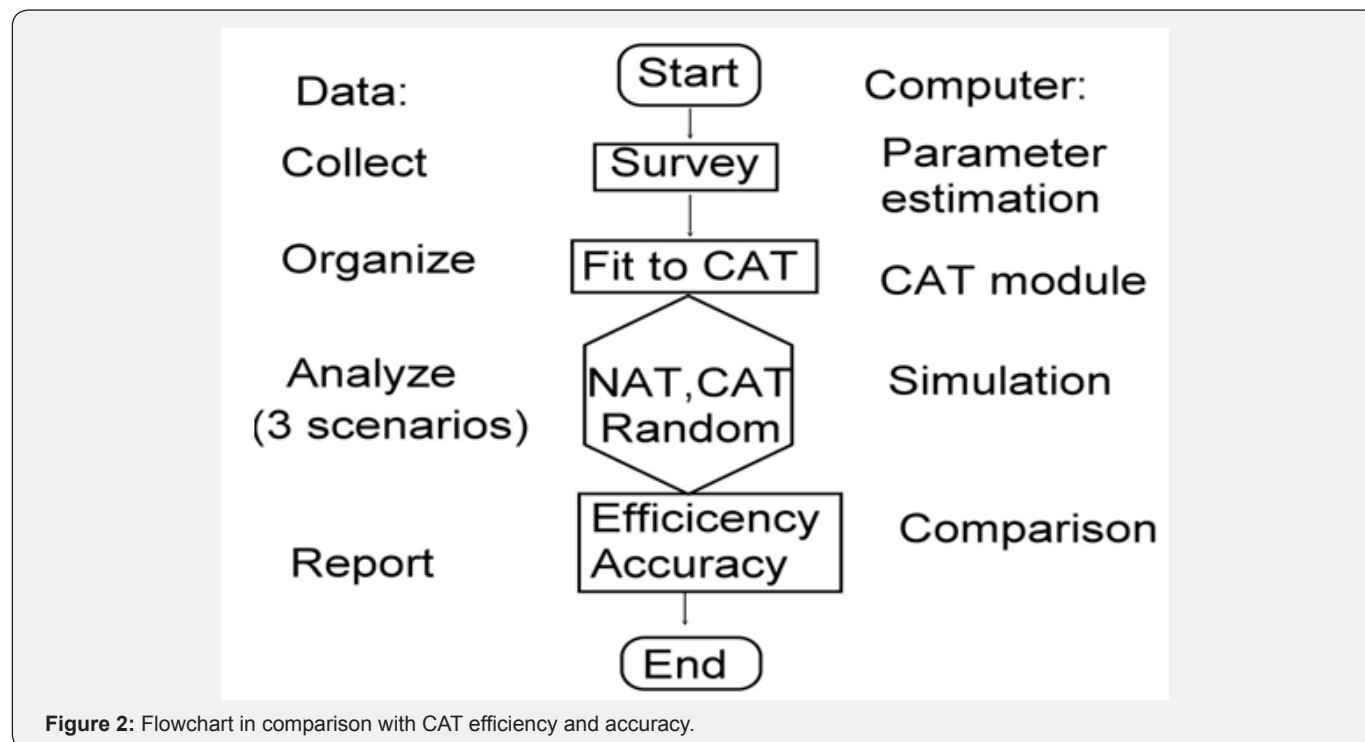


Figure 2: Flowchart in comparison with CAT efficiency and accuracy.

Three scenarios (i.e., NAT, CAT, and the randomly-selected method to NAT) were manipulated to compare their response efficiency and precision by comparing item length for answering questions, person measure; correlation coefficients; Smith's paired t tests [41]; estimated standard errors (SE) between CAT and the random to its counterpart of NAT (Figure 2) and Additional file 4. We ran an author-programmed VBA (Visual Basic for Applications) module in Microsoft Excel. Rasch person separation reliability of the TDS yielded by Winsteps (i.e., excluding all extreme scores summed to zero) was used to determine the CAT termination criterion using the standard error of measurement ($SEM=SD - reliability$). Another termination criterion is the mean of the last five change differences between the pre- and post estimated abilities on each CAT <0.05 . The minimum number of questions required for completion was set at 7 (7/22 items on TDS item length = 30%). The first item was randomly selected from the 22 items when starting the CAT. The provisional measures were estimated by the maximum log

likelihood estimation (MLE). The next question selected was the one with the most information obtained from the remaining unanswered items, interacting with the previously provisional person measures.

An online CAT was designed for smart phones

An online CAT was designed for examinees to report their depression scores in a unit of logit. The 22 items with their threshold difficulties (calibrated by Rasch Winsteps) and their responsive audios and pictures were uploaded to the website. The rules of the first and the next selected CAT item and the termination criteria are like the aforementioned simulation method.

Statistical tools and data analyses

SPSS 15.0 for Windows (SPSS Inc., Chicago, IL) and Med Calc 9.5.0.0 for Windows (Med Calc Software, Mariakerke, Belgium) were used to calculate (1) Cronbach's α , (2)

dimension coefficients [42], and (3) correlation coefficients between estimated person measures for CAT and the random to its counterpart of NAT. Independent t tests were used to compare (4) the ratios of the different paired person measures. RaschWinsteps was used for producing (5) person separation reliability. The prevalence (or incidence) rate is calculated by the formula (= the number of the depression grade excluded from the low stratum divided by the sample).

Results

The sample of 413 persons was obtained from the study (Additional file 3). Count distribution for the two study samples is shown in Figure 2.

Dimensionality

The TDS can be unidimensional because

- One factor was extracted using parallel analysis;
- All Infit and Outfit mean squares for the 22 items are in a range of 0.5 to 1.5 (the Infit column in Table 1; Figure 4).
- Item loadings from the Rasch PCA of residuals on the first contrast are standardized i.e., (loading – mean)/SD within –1.66 and 2.24 (<2.58, P>.01) in Table 1; PTME (point-measure) are between 0.71 and 0.84 (in the PTME column in Table 1) indicating high item loading to the unobserved latent trait.
- Rasch person separation reliability = 0.85, Cronbach α = 0.97, DC = 0.88 (> 0.67), and Smith's t test of proportions [41] is near to zero (= 1.4% = 11/414) outside the range +/- 1.96. In addition, category structure for the TDS displays the monotonically increasing threshold (-1.08, -0.52, and 1.60 logits) in compliance with Linacre's guidelines [40].

Table 1: One factor extracted from the scale with mean square between 0.50 and 1.50.

Item	Feature Type*	Difficulty Delta	SE	PTME	MNSQ Infit	PCA Z*	Loading
Intention to commit suicide	1	1.64	0.12	0.71	0.94	0.08	0.03
A poor appetite (gluttony)	3	1.01	0.11	0.73	1.09	0.23	0.07
A sense of guilt	1	0.91	0.11	0.77	0.93	1.18	0.32
Being reserved/ reticent/shy	4	0.88	0.11	0.76	0.96	-1.40	-0.36
Being useless	1	0.81	0.11	0.79	0.82	0.73	0.2
Being weepy	2	0.68	0.11	0.79	0.79	0.27	0.08
Lose interest in everything	1	0.61	0.11	0.77	1.01	0.92	0.25
Being scared/ frightened	2	0.55	0.11	0.74	1.18	-0.18	-0.04
Being easy to get angry	2	0.47	0.11	0.78	0.94	-0.07	-0.01
Being introverted (less talk)	4	0.46	0.11	0.77	1.04	-0.14	-0.03
A poor memory	3	0.05	0.11	0.81	0.89	1.49	0.40
More prefer to stay at home	4	-0.02	0.11	0.75	1.28	-0.79	-0.20
A smaller life circle/ few new things or friends to meet	4	-0.02	0.11	0.79	1.02	0.16	0.05
A sense of impotence/not knowing what to do	1	-0.14	0.11	0.84	0.78	0.65	0.18
Feeling ill	3	-0.29	0.10	0.81	1.00	-0.75	-0.19
Consideration with negative thoughts	1	-0.45	0.10	0.82	0.93	0.39	0.11
Being distracted/ deconcentrated	3	-0.47	0.10	0.83	0.83	0.04	0.02

Being depressed	2	-0.76	0.10	0.79	1.06	-1.02	-0.26
Poor sleep	3	-1.05	0.10	0.82	0.92	-1.66	-0.43
Being stressed	2	-1.46	0.10	0.79	1.07	-1.66	-0.43
Being exhausted	3	-1.62	0.10	0.80	1.01	-0.71	-0.18
Being worried/ anxious/upset	2	-1.78	0.10	0.80	0.97	2.24	0.60
Minimum		-1.78	0.10	0.71	0.78	-1.66	-0.43
Maximum		1.64	0.12	0.84	1.28	2.24	0.60

Note: Threshold difficulties are -1.08,-0.52, 1.60

*Type: 1.Recognition; 2.Emotion; 3.Physical status; 4.Person relation

Z = (loading –mean)/SD

Cutting point determination

The person separation reliability for the TDS is 0.85, indicating that three strata can be separated with thresholds at -0.7 and 0.7 with a highest Kappa coefficient and hit ratio (Table

2). The incidence rate of MDD for this study sample is 52.7% (= 218/413), Figure 2. We can see that three equal sizes with an equivalent accumulative probability are separated by the cutting points at -0.7 and 0.7, (Figure 3,4).

Table 2: Different cutting points yield validation results.

Cutting points

Cutting points	Group		Total	A Sensitivity	B Sensitivity	A+B Sum	LR+	LR-	Hit Ratio	Kappa
	Positive	Negative								
35(0.24 logits)										
>=35	146	0	146	0.69	1	1.69	-	0.31	0.84	0.68
<35	67	200	267							
27 (-0.43 logits)										
>=27	183	2	185	0.86	0.99	1.85	85.92	0.14	0.92	0.86
<27	30	198	228							
22 (-0.70 logits)										
>=24	194	8	202	0.91	0.96	1.87	22.76	0.09	0.93	0.87
<24	19	192	211							
18 (-1.29 logits)										
>=18	208	29	237	0.98	0.86	1.83	6.73	0.03	0.92	0.84
<18	5	171	176							
10(-2.27 logits)										
>=10	212	126	338	1	0.37	1.37	1.58	0.01	0.69	0.69
<10	1	74	75							

Note: *Hit ratio (accurate classification rate) = the number of accurate classification in both positive negative and cells divided by the total number.

LR+=Likelihood ratio positive; LR-=Likelihood ratio negative.

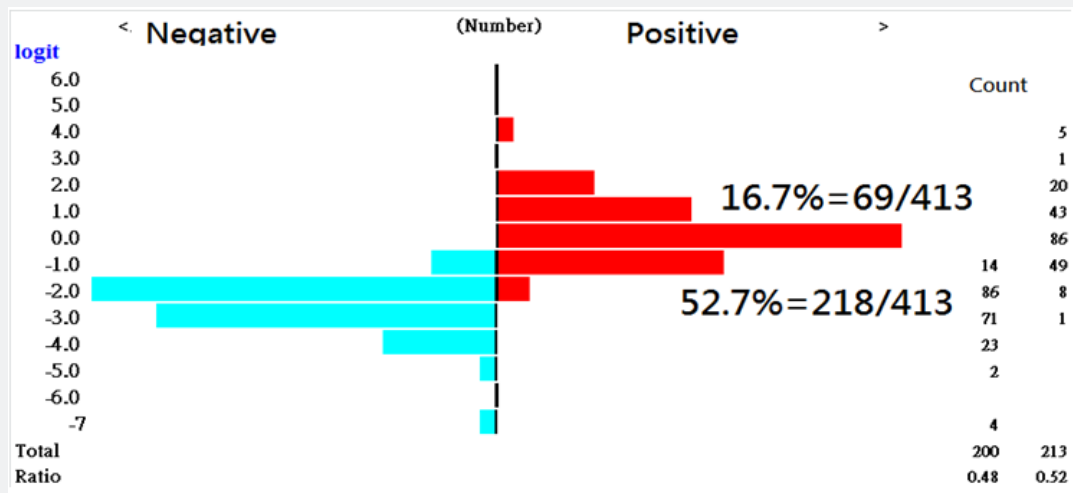


Figure 3: Count distribution for the two study samples.

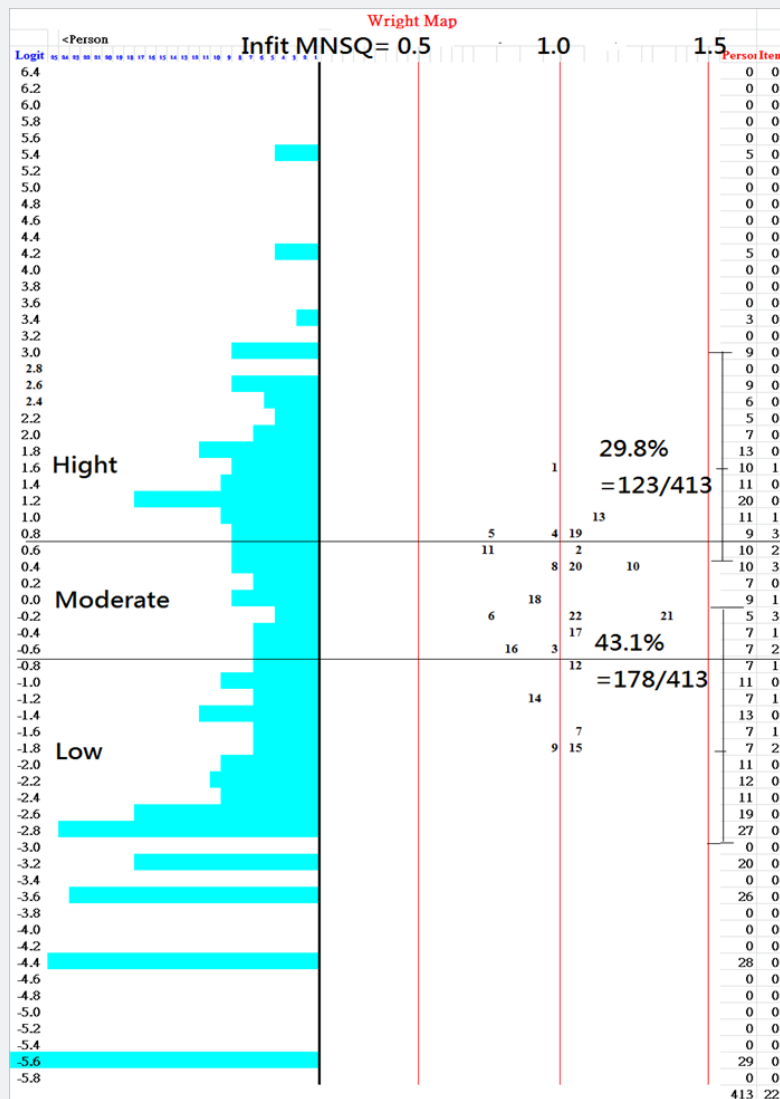


Figure 4: Item and person dispersion on an interval logit continuum scale.

Comparison of efficiency and precision

The CAT required substantially fewer items (mean = 14.3; SD = 0.39; SE = 0.28; 95% CI = 13.7-14.8, $p < .05$) than did NAT (= 22) and provided an efficient gain in test length of 36% (= 1-14/22), Figure 5 in panel A. Person measures from CAT did

not statistically differ from NAT because (1) Smith's t test of proportions [41] is 3.1% (= 13/413 < 5%), Figure 5 in panel B, and (2) correlation coefficient = 0.97 (= -square0.95, see Figure 5 in panel C). As compared to the random scenario, CAT earns a set of smaller SE, Figure 5 in panel D.

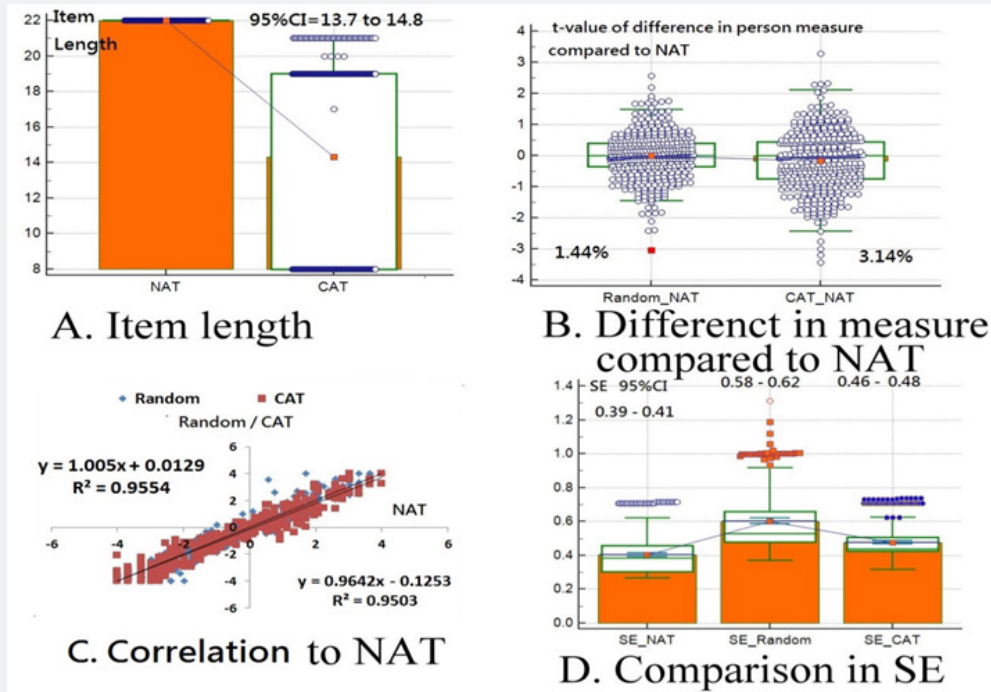


Figure 5: Comparison in efficiency and accuracy among scenarios.

Online NAQ-R assessment

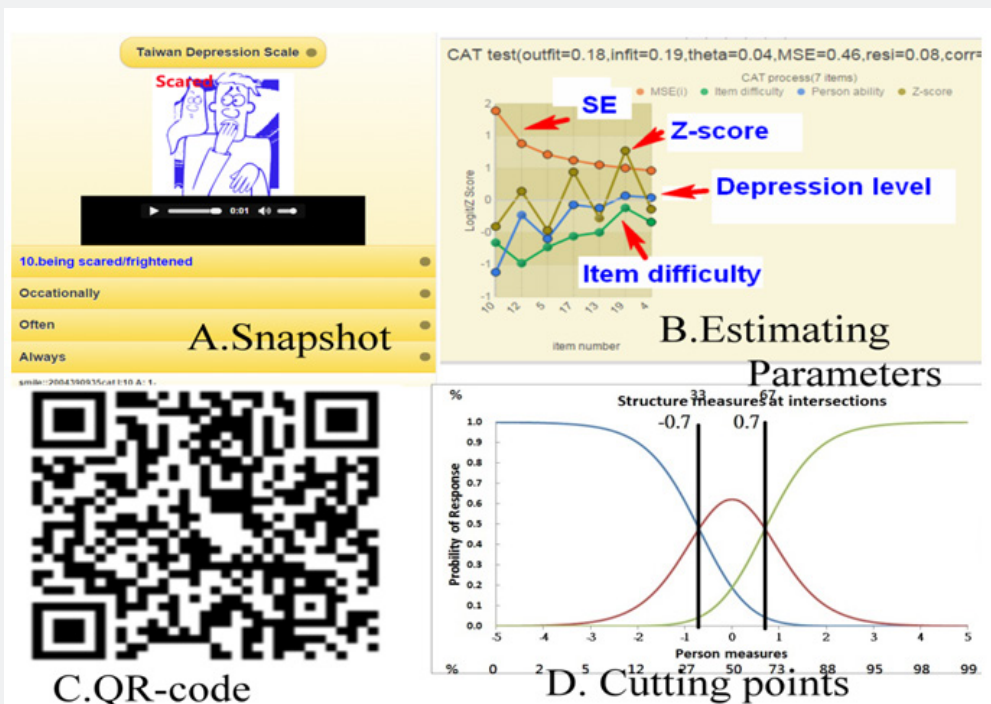


Figure 6: snapshot of online CAT-based Depression assessment.

By scanning a QR-code (Figure 6) at right bottom, the TDS item appears on the smartphone. We developed an online CAT module to demonstrate the assessment in action. The CAT processed each person item-by-item with picture animations (Figure 6) at left top. Adaptive item selection is based on maximizing information across unanswered items. The measurement of standard error (MSE) for each subscale decreased when the number of the items increased (Figure 6). The result with a person measure and the depression grade (i.e., low, moderate, or high) instantly shows on smart phone (Figure 6).

Discussion

Key findings

The results from this study indicate that the 22-item NAQ-R is unidimensional. A set of cutting point at -0.7 and 0.7 logits were determined for future use in workplace depression surveys. The incidence rate of depression for the study sample was 52.7%. The CAT is 36% more efficient for answering questions and achieved similar precision in measurements as did NAT. An available-for-download online CATNAQ-R APP for nurses was suited for smart phones.

What this adds to what was known

Consistent with the literature [43-48], the 22-item TDS can be unidimensional. The efficiency of CAT over NAT was supported. We confirm that CAT-based TDS requires significantly fewer answered items to measure depression symptom than NAT without compromising its measurement precision.

What it implies and what should be changed?

Cutoff point recommended for calculating depression prevalence rate

Many kinds of depression scale encounter a common problem that is to determine cutting points for calculating a depression prevalence rate. Different item length and category number lead to different cutting point if summation score is applied. In this study we determine cutting points at -0.7 and 0.7 that can be suitable for CAT in correspondence to different item length and can be referred to any kind of depression scale with different summation score using the percentage score (Figure 1).

For instance, a 20-item depression scale with 5 rating categories has two cutting points at $< 26 (= 33\% \times 80)$ and $< 52 (= 67\% \times 80)$, where 80 is the summation score ($= 20 \times 4$). Through which, a comparison between derived score levels and the suggested best cutoff points can help clinicians (or practitioners) assess examinees at risk of an incidence [19,20]. Multiple cutoff points are usually more powerful and useful than one single cutoff point [21,22]. Maslach et al. [49] suggested setting an equal sample size in each stratum as a way to determine cutting points. The value of 0.7 logit is the measurement of standard error beyond the mean of the sample. In this study the person SD=1.81, which is similar to the 1.7

adjustment for IRT because the person logistic ogive distribution (in logit units) is wider (i.e., 1.7 times) than the one with normal ogive distribution, see the difference in logit and probit [50].

At the end of 2016, more than 10,977 papers were found in a search with keyword "cut point". None discussed the determination of cutting points used for CAT with different item length for a respondent. In practice, we usually do not know the patient's true- and false-positive disease-specific status, like the TDS. The issue we face in clinical settings is how to identify the degree of patient incident problems. Through this study, if cutting points at -0.7 and 0.7 logits are selected for the TDS, the raw score in cutting points can be transformed by the formula ($= \text{total score} \times \text{the probability at 0.33 and 0.67}$), whereas 0.33 comes from the equation $\exp(-0.7)/(1 + \exp(-0.7))$ and 0.67 is from the equation $1 - \exp(-0.7)/(1 + \exp(-0.7))$, total score = 66 when 4-point (from 0 to 4) 22-item TDS is defined In Methods. The cutting points in raw score can be set at $< 22 (= 66 \times 0.33)$, and $\geq 44 (= 66 \times 0.67)$ to separate three strata in depression degree. The prevalence (or incidence) rate is easy to calculated and compared either with paper-and-pen format or with CAT in future.

Online CAT assessment

At the end of 2016, 757 papers were collected in US National Library of Medicine National Institutes of Health (pubmed.org) when searching keywords: computer adaptive testing. None was applicable using an online assessment suited for smart phones until the online skin cancer CAT was published [51]. We do ensure that more papers in future will be published on the usefulness of online CAT as with all forms of Web-based technology are rapidly increasing [52].

Unidimensional scale detection

Many studies [42,53] reported the issue of scale unidimensionality detection. From the Library of PubMed and BioMed Central, we got 1,005 and 333 papers with the keyword "unidimensionality", 359,957 and 23,902 results for "depression". In the current study, we demonstrated the method Tennant [54] suggested using three steps to assess scale unidimensionality: Conduct prior testing using Horn's parallel analysis; use Rasch fit statistics; run post hoc tests using Rasch standardized residual loading, and Smith [41] independent t-tests to compare estimates of the percentages ($< 5\%$, within ± 1.96). In addition, the dimension coefficient (≥ 0.67 , DC) and PTME (> 0.40) included in detecting scale unidimensionality are recommended to readers.

Strengths of this study

Four goals have been reached in this study: We demonstrate a Taiwanese Depression Scale (TDS) that is a unidimensional construct (2) cutting points at -0.7 and 0.7 logits were recommended to future studies in computing depression prevalence rate at work place using TDS; CAT gains 36% efficient than did NAT, and; online CAT is applicable in practice. Among

them, the reason for 36% efficient than did NAT is because we added another termination rule in CAT: the mean of the last five change differences between the pre-and post estimated abilities on each CAT less than 0.05. Through the termination rule of detecting the last five change differences in estimated abilities less than 0.05 makes the item length less than that in other studies [42,53]. If all CAT cases are controlled by the only termination rule of SE less than 0.44 ($= \text{SQRT}(1 - 0.8) = \text{SQRT}(1 - \text{reliability})$), the precision measured by SE on CAT (Figure 3) will be substantially higher than the dual stop conditions we did in this study because a longer item length leads to a high reliability (or a smaller measure SE) than a shorter one.

In addition, the online CAT with audio and picture animations is available for interested readers to practice if scanned on the QR-code in Figure 6 which is rare in any previously published articles. Furthermore, cutting points set at -0.7 and 0.7 logits with an equal stratum member size might be generalized to other incidences or diseases when the patient's true- and false-positive disease-specific status is not known beforehand. Like the TDS, we merely intend to identify the grade of the incidence and compare to the norm.

Limitations of the study

Several issues should be considered more thoroughly in further. First, the secondary data source limits us not to identify differential item functioning (DIF) on gender or other race groups. Second, the high incidence rate (52.7%) cannot be generalized to the prevalence rate because the sample (comprising 213 depression patients and 200 normal undergraduates) was particularly manipulated for verifying the TDS validation only (Table 1) instead of calculating prevalence rate in a real world. More studies are recommended to assess the generalizability of the study with different samples using the same cutting points and the same version of TDS in future. Third, the online CAT is not equipped with much useful functionality as we expected in practice. Such as protecting cheating behaviors and detecting aberrant responses that are required to be in future advanced versions. Fourth, although the scale's Cronbach's α coefficients was 0.96, we conservatively determined that the scales' person strata were three according to Rasch separation reliability = 0.85 and literature [36-38]. Multiple cutoff points are not limited to three strata if the separation index reaches an extremely higher level, which will affect the determination of appropriate cutting points of the TDS.

Conclusion

The CAT-based TDS forming a unidimensional construct reduces respondents' burden without compromising measurement precision and increases endorsement efficiency. The online TDS module developed by the authors is recommended for assessing hospital employees or other workplace members using the criteria at -0.7 and 0.7 (or <22 and <44 in summed score) to identify depression grade as one of the three levels (high, moderate, and low).

Declaration

Ethics approval and consent to participate

The secondary data were retrieved from two published papers [31,32] both for CAT used as item pool and for simulation as well as for demonstration in a MS Excel format. The way we extracted data from papers is fully disclosed with a video in Additional file 3.

Availability of data and materials

All data used for verifying the proposed computer module during this study are extracted from two published papers [31,32]. The Microsoft Excel-based computer module including the demonstrated data can be downloaded from the supplementary information files.

Authors contribution

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.

Acknowledgement

We thank Frank Bill who provided medical writing services to the manuscript: Additional files

Additional File 1

Data included in the Winsteps control file

Additional File 2

Study data and Table 1 saved and organized in a MS Excel format

Additional File 3

How to extract data from published papers

<http://www.healthup.org.tw/marketing/course/information/getdatafrompaper.mp4>

Additional File 4

Introduction to CAT simulation process and comparisons with results

<http://www.healthup.org.tw/marketing/course/information/CATsimulation.mp4>

Additional File 5

Demonstration of an online CAT using the TDS tool

<http://www.healthup.org.tw/marketing/course/information/DepressionCATonline.mp4>

References

- Hidaka BH (2012) Depression as a disease of modernity: explanations for increasing prevalence. *J Affect Disord* 140(3): 205-214.
- (2013) National Institute of Mental Health [NIMH]. The numbers count: Mental disorders in America.
- Kessler RC, Berglund P, Demler O, Jin R, Koretz D, et al. (2003) The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 289(23): 3095-3105.
- The World Health Organization (2008) The global burden of disease: 2004 update, Table A2: Burden of disease in DALYs by cause, sex and income group in WHO regions, estimates for 2004. Geneva, Switzerland.
- Kessler RC, Chiu WT, Demler O, Walters EE (2005) Prevalence, severity and comorbidity of twelve-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). *Archives of General Psychiatry* 62(6): 617-627.
- (2001) US Census Bureau Population Estimates by Demographic Characteristics: Annual Estimates of the Population by Selected Age Groups and Sex for the United States.
- Kessler RC, Berglund PA, Demler O, Jin R, Walters EE (2005) Life time prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). *Archives of General Psychiatry* 62(6): 593-602.
- Eaton WW, Anthony JC, Gallo J (1997) Natural history of diagnostic interview schedule/DSM-IV major depression. The Baltimore Epidemiologic Catchment Area follow-up. *Arch Gen Psychiatry* 54(11): 993-999.
- Kessler RC, Bromet EJ (2013) The epidemiology of depression across cultures. *Annu Rev Public Health* 34: 119-138.
- Hwu HG, Chang IH, Yeh EK, Chang CJ, Yeh LL (1996) Major depressive disorder in Taiwan defined by the Chinese diagnostic Interview Schedule. *J Nerv Ment Dis* 184(8): 497-502.
- Liao SC, Chen WJ, Lee MB, Lung FW, Lai TJ, et al. (2012) Low prevalence of major depressive disorder in Taiwanese adults: possible explanations and implications. *Psychol Med* 42(6): 1227-1237.
- Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, et al. (1996) Cross-national epidemiology of major depression and bipolar disorder. *JAMA* 276(4): 293-299.
- Zung WW (1965) A Self-Rating Depression Scale. *Arch Gen Psychiatry* 12: 63-70.
- Hamilton M (1967) Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol* 6(4): 278-296.
- Williams JB (1988) A structured interview guide for the Hamilton Depression Rating Scale. *Arch Gen Psychiatry* 45(8): 742-747.
- Montgomery SA, Asberg M (1979) A new depression scale designed to be sensitive to change. *Br J Psychiatry* 134(4): 382-389.
- Williams JBW, Kobak KA (2008) Development and reliability of a structured interview guide for the Montgomery-Asberg Depression Rating Scale (SIGMA). *The Br J Psychiatry* 192(1): 52-58.
- Hwang AW, Chou YT, Hsieh CL, Hsieh WS, Liao HF (2015) A developmental screening tool for toddlers with multiple domains based on Rasch analysis. *J Formos Med Assoc* 114(1): 23-34.
- Chien TW, Lin WS (2016) Simulation study of activities of daily living functions using online computerized adaptive testing. *BMC Med Inform Decis Mak* 16(1): 130.
- Straus E, Richardson WS, Glaszow P, Haynes RB (2005) Evidence-based medicine: how to practice and teach EBM (3rd edn.). Elsevier Churchill Livingstone, London.
- Liao HF, Yao G, Chienc CC, Cheng LY, Hsieh WS (2014) Likelihood ratios of multiple cutoff points of the Taipei City Developmental Checklist for Preschoolers, 2nd version. *Form J Med* 113(3): 179-186.
- Chien TW, Wu HM, Wang WC, Castillo RV, Chou W (2009) Reduction in patient burdens with graphical computerized adaptive testing on the ADL scale: Tool development and simulation. *Health Qual Life Outcomes* 7: 39.
- Chien TW, Wang WC, Huang SY, Lai WP, Chou JC (2011) A web-based computerized adaptive testing (CAT) to assess patient perception of hospitalization. *J Med Internet Res* 13(3): e61.
- Ma SC, Chien TW, Wang HH, Li YC, Yui MS (2014) Applying computerized adaptive testing to the negative acts questionnaire-revised: Rasch analysis of workplace bullying. *J Med Internet Res* 16(2): e50.
- De Beurs DP, De Vries AL, De Groot MH, De Keijser J, Kerkhof AJ (2014) Applying computer adaptive testing to optimize online assessment of suicidal behavior: A simulation study. *J Med Internet Res* 16(9): e207.
- Stochl J, Böhneke JR, Pickett KE, Croudace TJ (2016) 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 Medical Research Methodology* 16: 58.
- Eack SM, Singer JB, Greeno CG (2008) Screening for anxiety and depression in community mental health: the beck anxiety and depression inventories. *Community Ment Health J* 44(6): 465-474.
- Ramirez BM, Bostic JQ, Davies D, Rush AJ, Witte B, et al. (2000) Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry* 157(10): 1599-1605.
- Shear MK, Greeno C, Kang J, Ludewig D, Frank E, et al. (2000) Diagnosis of nonpsychotic patients in community clinics. *Am J Psychiatry* 157(4): 581-587.
- Yu MN, Liu YJ, Li RH (2008) The practical usage of cutoff score in the Taiwanese depression scale. *Journal of Educational Research and Development* 4(4): 231-258.
- Yu MN, Huang HY, Liu YJ (2011) The development and psychometric study of Taiwan depression scale. *Psychological Testing* 58(3): 479-500.
- Rasch G (1980) Probabilistic models for some intelligence and achievement test. Danish Institute for Educational Research Expanded. University of Chicago Press, Chicago, USA.
- Andrich D (1978) A rating formulation for ordered response categories. *Psychometrika* 43: 561-573.
- Linacre JM (2007) How to Simulate Rasch Data. *Rasch Measurement Transactions* 21(3): 1125.
- Fisher WJ (1994) Reliability, separation, strata statistics. *Rasch Meas Trans* 6(3): 238.
- Wright BD, Masters GN (2002) Number of person or item strata. *Rasch Meas Trans* 16(3): 888.
- Wright BD (1996) Reliability and separation. *Rasch Meas Trans* 9(4): 472.
- Fisher WP (2008) The cash value of reliability. *Rasch Meas Trans* 22(1): 1160-1163.
- Linacre JM (2003) Optimizing rating scale category effectiveness. *Journal of Applied Measurement* 3 (1): 85-106.
- Smith EV (2002) Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* 3(2): 205-231.

41. Chien TW (2012) Cronbach's Alpha with the Dimension Coefficient to Jointly Assess a Scale's Quality. *Rasch Measurement Transactions* 26: 3.
42. Ma SC, Chien TW, Wang HH, Li YC, Yui MS (2014) Applying computerized adaptive testing to the negative acts questionnaire-revised: Rasch analysis of workplace bullying. *J Med Internet Res* 16(2): e50.
43. Chien TW, Wang WC, Huang SY, Lai WP, Chow JC (2011) A web-based computerized adaptive testing (CAT) to assess patient perception in hospitalization. *J Med Internet Res* 13(3): e61.
44. Wainer HW, Dorans NJ (1990) *Computerized Adaptive Testing: A Primer*. Hillsdale, L Erlbaum Associates, New Jersey, USA.
45. Embretson S, Reise S, Reise SP (2000) *Item Response Theory for Psychologists*. Mahwah, New Jersey, USA.
46. Djaja N, Janda M, Olsen CM, Whiteman DC, Chien TW (2016) Estimating Skin Cancer Risk: Evaluating Mobile Computer-Adaptive Testing. *J Med Internet Res* 18(1): e22.
47. Maslach C, Schaufeli WB, Leiter MP (2001) Job burnout. *The Annual Review of Psychology* 52: 397-422.
48. Linacre JM (2016) Logit and Probit: what are they? *Winsteps guideline*.
49. Mitchel SJ, Godoy L, Shabazz K, Horn IB (2014) Internet and mobile technology use among urban African American parents: survey study of a clinical population. *J Med Internet Res* 16(1): e9.
50. Smith RM (1996) A Comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling* 3: 25-40.
51. Zwick WR, Velicer WF (1986) Comparison of the rules for determining the number of components to retain. *Psychological Bulletin* 99(3): 432-442.
52. Wright BD (1994) Unidimensionality coefficient. *Rasch Measurement Transactions* 8: 385.
53. Linacre JM (2011) Rasch Measures and Unidimensionality. *Rasch Measurement Transactions* 24(4): 1310.
54. Tennant A, Pallant JF (2006) Unidimensionality matters! (A tale of two Smiths?). *Rasch Meas Trans* 20(1): 1048-1051.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/PBSIJ.2018.08.555750](https://doi.org/10.19080/PBSIJ.2018.08.555750)

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
(Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
<https://juniperpublishers.com/online-submission.php>