UNIVERSITI TEKNOLOGI MARA
Cawangan Kedah
Kampus Sungai Petani

2021 iCMS
INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

e-PROCEEDINGS

of The 5th International Conference on Computing, Mathematics and Statistics (iCMS2021)

4-5 August 2021

Driving Research Towards Excellence

e-Proceedings of the 5th International Conference on Computing, Mathematics and Statistics (iCMS 2021)

Driving Research Towards Excellence

Editor-in-Chief: Norin Rahayu Shamsuddin

Editorial team:
Dr. Afida Ahamad
Dr. Norliana Mohd Najib
Dr. Nor Athirah Mohd Zin
Dr. Siti Nur Alwani Salleh
Kartini Kasim
Dr. Ida Normaya Mohd Nasir
Kamarul Ariffin Mansor

e-ISBN: 978-967-2948-12-4
DOI

Library of Congress Control Number:

Publication by
Department of Mathematical Sciences
Faculty of Computer & Mathematical Sciences
UiTM Kedah

TABLE OF CONTENT

PART 1: MATHEMATICS

PART 2: STATISTICS

PART 3: COMPUTER SCIENCE & INFORMATION TECHNOLOGY

PART 4: OTHERS

DETECTION OF DIFFERENTIAL ITEM FUNCTIONING FOR THE NINE-QUESTIONS DEPRESSION RATING SCALE FOR THAI NORTH DIALECT

Suttipong Kawilapat[1], Benchlak Maneeton[2], Narong Maneeton[3], Sukon Prasitwattanaseree[4], Thoranin Kongsuk[5], Suwanna Arunpongpaisal[6], Jintana Leejongpermpool[7], Supattra Sukhawaha[8], Patrinee Traisathit[9]

[1] Department of Statistics, Faculty of Science, Chiang Mai University, Thailand; Department of Psychiatry, Faculty of Medicine, Chiang Mai University, Thailand, [2,3] Department of Psychiatry, Faculty of Medicine, Chiang Mai University, Thailand, [4] Department of Statistics, Faculty of Science, Chiang Mai University, Thailand, [5] Phra Si Maha Phot Psychiatric Hospital, Ubon Ratchathani, Thailand; Somdet Chaopraya Institute of Psychiatry, Bangkok, Thailand, [6] Department of Psychiatry, Faculty of Medicine, Khon Kaen University, Thailand, [7,8] Phra Si Maha Phot Psychiatric Hospital, Ubon Ratchathani, Thailand, [9] Department of Statistics, Faculty of Science, Chiang Mai University, Thailand; Data Science Research Center, Department of Statistics, Faculty of Science, Chiang Mai University, Thailand

([1] suttipong.kawilapat@gmail.com, [2] benchalak.maneeton@cmu.ac.th, [3] narong.m@cmu.ac.th, [4] sprasitwattanaseree@gmail.com, [5] tkongsuk@gmail.com, [6] suwaru@kku.ac.th, [7] virgojinny12@gmail.com, [8] s_sukawa@yahoo.com, [9] patrinee.t@cmu.ac.th)

A different probability of response to an item among respondents with the same underlying trait, differential item functioning (DIF), may compromised comparisons across subgroups and lead to misleading results. This cross-sectional study to detect the DIF among 1,475 participants who participated in a validity study of the Nine-Questions Depression Rating Scale (9Q) for Thai North dialect. The 9Q was consisted of 9 polytomous items assessed for the severity and frequency of depression symptoms according to DSM-5 within 2 weeks. The detection of DIF was considered between gender, age, underlying disease, and income using item response theory (IRT)-based and ordinal logistic regression (OLR)-based approach. The IRT difficulty parameters for each item and threshold parameters for each possible category were estimated based on the generalized partial credit model. Item 1 has the highest discrimination parameter following by item 9 (2.533 and 2.387, respectively). The IRT-based approach seemed to be proper than OLR-based approach in the examination of DIF and scoring due to the different of discrimination parameters. The DIF were presented between gender (item 5 and 8), age (item 1, 2, 4, and 7), and underlying disease (item 2, 3, 5, 6, and 7). Scoring accounted for the discrimination and threshold parameters based on IRT approach might be useful to reduce the bias of depression assessment.

***Keywords*:** Depression, Nine-Questions Depression Rating Scale, Differential item functioning, Item response theory, Thai

1. Introduction

Depression is a common mental disorder that is a leading cause of the global disease burden and suicide. In 2017, an estimated 264 million people (3.44%; range 2–6%) worldwide and 2.62 million people (3.09%) in Thailand experienced depression. The prevalence of depression in Thailand is slightly different between males and females (2.57% vs. 3.56%) and around twice higher in the elderly (50 years of age or more) than individuals aged 15–49 years old (6.02–6.29% vs. 3.37%) (Ritchie and Roser, 2018).

The measurement of psychological traits, such as depression and the quality of life, is complicated due to there being no measurement to assess those traits directly. However, it can be quantified with an instrument, of which there are several for depression assessment, such as the Hamilton Rating Scale for Depression, the Beck Depression Inventory, the Montgomery-Åsberg Depression-Rating Scale, and the Patient Health Questionnaire-9. A Nine-Questions Depression Rating Scale (9Q) for the Thai North dialect is one of the measurement tools developed to use for the assessment of depression severity in Thailand. It consists of nine rating scale items about the frequency and severity of the diagnostic symptom criteria for depression (Kongsuk et al., 2018).

The observed scores during measurement can be different between groups due to differences in the true trait ability or differences among the characteristics of the respondents. Different probabilities of response to an item among respondents with the same underlying trait score is defined as differential item functioning (DIF), the presence of which may compromise comparisons across subgroups and can lead to misleading results (Crane et al., 2004). The traditional calculation of the measurement score when not considering DIF might not be proper when the latter occurs, and so the aim of this study is to determine the presence of DIF in responses to depression assessment tools and related factors pertaining to the Thai population.

2. Methodology

2.1 Setting and Participants

We used secondary data from a study on the criterion-related validity of a revised 9Q in the northern Thai dialect comprising 1,527 individuals from the northern region of Thailand. This revised questionnaire was translated from the central Thai dialect version. Participants who did not complete all items in the assessment were excluded from the study. Prior to the analyses for the presence of DIF, the demographic characteristics of the remaining participants were obtained: gender, age group (adolescent: less than 18 years old, adult: 18–59 years old, and elderly: 60 years or higher), underlying diseases, and income (≤5,000, 50,000–10,000, and >10,000 baht per month).

2.2 Depression Assessment

The 9Q was used as a measurement of depression in this study. It consists of nine rating scale items about the severity (0 = no symptoms, 1 = mild, 2 = moderate, 3 = severe) and frequency (1 = several days, 2 = more than 7 days, 3 = nearly every day) of depression symptoms within the previous two weeks according to the fifth edition of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013). The score of each item is calculated as the product of the severity and frequency scores. The total score for the 9Q ranges from 0 to 81 points.

2.3 Determining the Presence of DIF

DIF occurs when participants from different demographic groups (e.g., gender, age) with the same underlying trait score have a different probability of responding to an item. It can either be non-uniform, which is due to a statistically significant interaction between the trait level and the demographic variable (effect modification), or uniform, which is the difference between the strength of the relationship between the ability and item responses in a model with and without the demographic variable for each item (confounding) (Crane et al., 2006).

The ordinal logistic regression (OLR) technique is an approach based on observing variables to determine DIF for polytomous items. The following ordinal response models are fitted prior to exploring both uniform DIF (UDIF) and non-uniform DIF (NUDIF):

$$f\,(\text{item response}) = ß_0 + ß_1\theta, \quad (1)$$

$$f\,(\text{item response}) = ß_0 + ß_1\theta + ß_2X, \quad (2)$$

$$f\,(\text{item response}) = ß_0 + ß_1\theta + ß_2X + ß_3(\theta X), \quad (3)$$

where $ß_0$ are the intercept coefficients, $ß_1$ are the regression coefficients of the trait level θ, $ß_2$ are the regression coefficients of variable X, and $ß_3$ are the regression coefficients of the interaction between trait level θ and variable X.

For NUDIF, the difference between the -2 log-likelihood values of the models in (2) and (3) is compared with a Chi-squared distribution (1 degree of freedom), while NUDIF is defined as p-value <0.05. For UDIF, the relative difference between the coefficients of the trait level ($ß_1$) between the models in (1) and (2) is determined. If the relative difference is at least 10% or there is significance in the likelihood between the models, then UDIF has occurred.

Since the OLR approach is considered to be the DIF according to the sum score, which rarely signifies a sufficiently precise measure of underlying trait, an approach based on the item response theory (IRT) is proposed as an alternative method to determine DIF. The baseline IRT models are fitted for all items and then compared to the other model with varied discrimination and threshold parameters between the reference and focal groups for each item. A comparison of models is performed using the likelihood ratio test, with a significant difference (p-value <0.05) between the baseline and constrained model indicating the presence of DIF between the groups (Raykov and Marcoulides, 2018).

2.4 Statistical Analysis

The demographics of the participants are reported as frequencies and percentages for categorical variables and as medians and interquartile ranges (IQRs) for the continuous variables. The OLR approach was performed using the Stata "DIFDETECT" command (Crane et al., 2006). The generalized partial credit model (GPCM) (4), which is an IRT model for polytomous items, was preferable for estimating the IRT parameters in this study (Muraki, 1992; Edelen and Reeve, 2007; Nering and Ostini, 2010):

$$P_{ik}(\theta) = \frac{\exp\left[\sum_{k=1}^{m} a_i\left(\theta - b_{ik}\right)\right]}{\sum_{h}^{m-1}\exp\left[\sum_{k=1}^{h} a_i\left(\theta - b_{ik}\right)\right]}, \tag{4}$$

where $P_{ik}(\theta)$ is the probability of responding to item i in category k (k = 0, 1, … , m), a_i is the discrimination parameter of item i, and b_{ik} is the threshold parameter for item i in category k.

All analyses were performed using Stata 17 (StataCorp, College Station, Texas, USA).

3. Results and Discussions

Of 1,475 participants included in the study, the majority were female (66.6%), adult (73.6%), did not have any underlying diseases (63.8%), and had an income per month of less than 5,000 baht. The median age of the participants was 45 (IQR 33–57) years old. Some participants did not complete their information about underlying disease and income on patient report form (Table 1).

Table 1: Demographic characteristics of the participants (N=1,475).

Characteristic	n (%) or Median [IQR]
Gender	
Male	493 (33.4%)
Female	982 (66.6%)
Age	45 [33–57]
13–18	120 (8.1%)
18–59	1,086 (73.6%)
60+	269 (18.3%)
Underlying disease (n = 1,459)	
No	931 (63.8%)
Yes	528 (36.2%)
Income (baht/month) (n = 1,453)	
<5,000	759 (52.2%)
5,000–10,000	442 (30.4%)
>10,000	252 (17.3%)

According to the item endorsement in Figure 1, more than 80% of the participants had no symptoms related to depression within the previous two weeks, except for items 2, 3, and 7. Item 3 had the highest endorsement rate of having severe symptoms nearly every day. Almost all of the participants (96%) did not report thoughts of physical self-harm or suicide (item 9). Table 2 reports the discrimination and threshold parameters for all possible combinations of categories for each item. The highest discrimination parameter value was obtained for item 1, followed by items 9, 2, and 4, all of which are the most related to depression. The discrimination parameter values for items 3 and 5 were lower than the others.

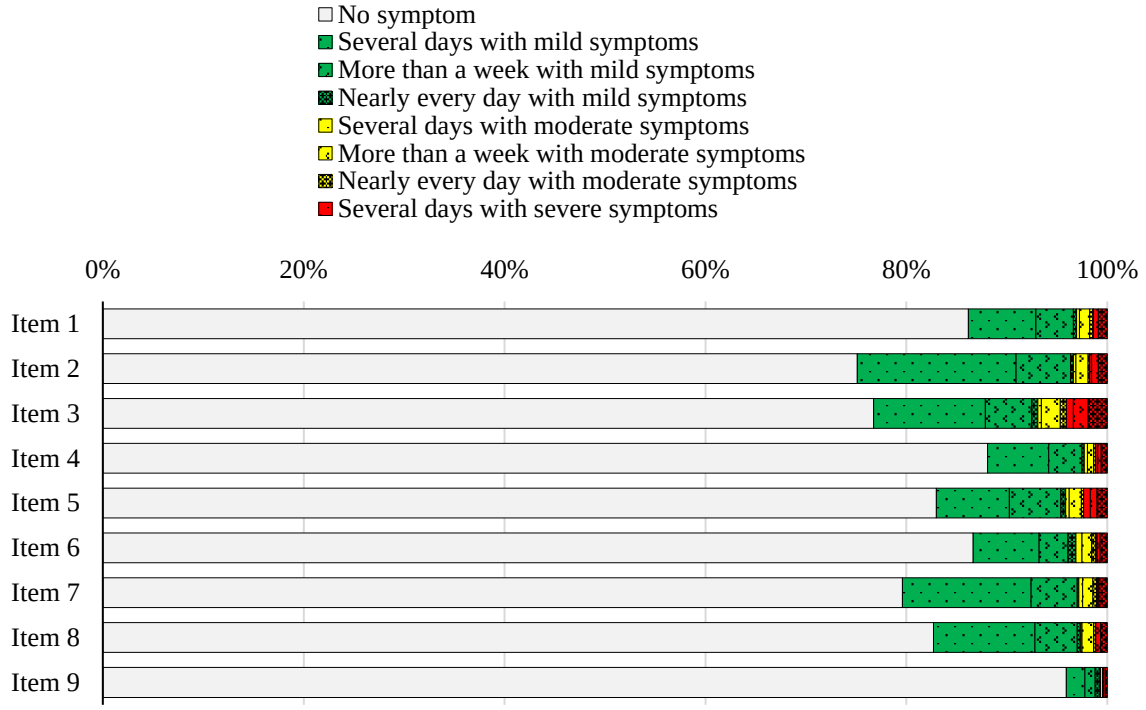

Figure 1: Item endorsement by the participants.

Table 2: Estimated parameter based on item response theory using graded response model.

Item	IRT Parameters from the GPCM									
	a_i	$b_{i(11)}$	$b_{i(12)}$	$b_{i(13)}$	$b_{i(21)}$	$b_{i(22)}$	$b_{i(23)}$	$b_{i(31)}$	$b_{i(32)}$	$b_{i(33)}$
1. Depressed mood	2.533	1.508	1.442	2.399	1.659	1.126	2.140	2.393	1.046	1.642
2. Markedly diminished interest or pleasure	1.885	1.069	1.549	2.910	1.489	0.793	2.829	1.509	1.145	1.713
3. Insomnia or hypersomnia	0.376	5.117	2.623	5.924	2.334	-3.624	4.126	0.952	-0.762	0.778
4. Fatigue or loss of energy	1.837	1.890	1.484	2.957	1.295	1.338	2.322	1.781	1.563	1.591
5. Significant weight loss or gain	0.428	5.734	1.188	6.247	1.701	-1.889	5.430	-1.448	1.456	0.587
6. Feeling of worthlessness or excessive or inappropriate guilt	1.167	2.473	1.617	2.318	1.681	1.174	2.363	2.649	1.178	0.971
7. Diminished ability to think or concentrate, or indecisiveness	1.276	1.650	1.695	4.019	0.590	0.851	2.514	2.441	1.787	0.450
8. Psychomotor agitation or retardation	1.287	1.889	1.608	3.236	2.181	-0.076	3.021	1.091	-	1.547
9. Recurrent thoughts of death, recurrent suicidal ideation, or a suicide attempt	2.387	2.372	1.679	1.962	2.564	1.367	2.070	1.979	-	1.905

Abbreviations: IRT, the item response theory; GPCM, the generalized partial credit model; a_i, the discrimination parameter of item i; $b_{i(fs)}$, the threshold parameter of item i with frequency f and severity s.

Table 3: Detecting the presence of DIF using the OLR approach.

	Gender			Age			Underlying Disease			Income		
	Non-uniform	Uniform		Non-uniform	Uniform		Non-uniform	Uniform		Non-uniform	Uniform	
Item	(*p*)	(%)	(*p*)	(*p*)	(%)	(*p*)	(*p*)	(%)	(*p*)	(*p*)	(%)	(*p*)
1	0.325	0.0001	0.073	0.995	0.0024	0.386	**0.009**	-0.0015	0.838	0.189	-0.0031	0.674
2	**0.016**	0.0060	**0.002**	0.071	0.0264	**<0.001**	**<0.001**	0.0262	**0.006**	0.861	0.0024	0.303
3	0.665	-0.0006	0.072	0.228	0.0014	**0.009**	0.217	-0.0231	**0.003**	0.067	-0.0040	**0.007**
4	0.071	-0.0003	0.895	**0.004**	0.0113	**0.002**	0.920	-0.0118	**0.024**	0.055	0.0004	**0.021**
5	0.257	0.0005	0.734	**0.023**	0.0010	0.919	**0.009**	-0.0169	0.121	0.574	-0.0014	0.148
6	0.829	0.0016	0.310	0.363	0.0254	**<0.001**	0.459	0.0146	0.215	0.194	-0.0026	0.359
7	0.299	0.0003	0.895	**0.010**	0.0110	**0.005**	0.165	0.0028	0.716	0.230	-0.0035	0.529
8	0.730	0.0065	0.842	0.422	-0.0008	0.779	**0.004**	-0.0199	**0.001**	0.416	-0.0042	0.684
9	0.141	-0.0005	0.837	0.564	-0.0047	0.847	0.431	-0.0240	0.068	0.410	0.0124	0.024

The items in bold indicate the presence of DIF (differential item functioning).

Table 4: Detecting the presence of DIF using the IRT approach.

	Gender		Age		Underlying disease		Income	
Item	Non-uniform (*p*)	Uniform (*p*)	Non-uniform (*p*)	Uniform (*p*)	Non-uniform (*p*)	Uniform (*p*)	Non-uniform (*p*)	Uniform (*p*)
Item 1	0.408	0.320	**0.018**	**0.016**	0.843	0.836	0.546	0.420
Item 2	0.094	0.066	**0.001**	**0.004**	**0.005**	**0.008**	0.241	0.229
Item 3	0.112	0.242	0.126	0.126	**0.049**	**0.031**	0.100	0.116
Item 4	0.687	0.586	**0.004**	**0.002**	0.599	0.520	0.775	0.734
Item 5	**0.003**	**0.017**	0.219	0.366	0.051	**0.034**	0.179	0.150
Item 6	0.314	0.240	0.351	0.261	0.240	0.196	0.194	0.185
Item 7	0.868	0.802	**0.035**	0.185	**0.041**	0.255	0.653	0.653
Item 8	**0.006**	**0.003**	0.161	0.086	**0.003**	**0.003**	0.084	0.125
Item 9	0.568	0.761	0.838	0.753	0.568	0.452	0.709	0.713

The items in bold indicate the presence of DIF (differential item functioning).

According to DIF detection using the OLR approach, DIF between the characteristics was present in all of the studied variables, including gender (item 2), age (items 2, 3, 4, 5, 6, and 7), underlying disease (items 1, 2, 3, 4, 5, and 8), and income (items 3 and 4). NUDIF was present across the groups for all of the studied variables except income, while UDIF was present across all of the characteristics (Table 3).

In the detection of DIF based on the IRT approach, DIF was present for gender (items 5 and 8), age group (items 1, 2, 4, and 7), and underlying disease (items 2, 3, 5, 7, and 8). These variables also contained both NUDIF and UDIF across the groups. However, DIF was not present for income group (Table 4).

The findings from the DIF analyses based on the OLR and IRT approaches were inconsistent. Items presenting with DIF were different between the approaches for some variables (i.e., age and underlying disease). The most inconsistent presence of DIF was found for age and underlying disease. DIF was significant for age (items 3, 5, and 6) using the OLR approach but not with the IRT approach, while DIF was only significant for item 1 in the IRT approach. The estimated IRT parameters in our study also showed various discrimination parameters in each 9Q item. Findings from previous studies suggest that the IRT technique may reveal additional information about the actual level of the underlying trait compared to the observed score (Reise and Haviland, 2005; Snitz et al., 2012; Gorter et al., 2015; Saracino et al., 2020). Since DIF occurs from the effect modification between the characteristics and the trait level or the difference in the strength of the relationship between the ability and item responses, the inconsistency in the OLR approach might have resulted from not accounting for the different discrimination and threshold parameters related to depression for each item. Therefore, the IRT approach is probably more appropriate for examining the DIF in polytomous items.

Both NUDIF and UDIF were present for gender in two items (“item 5: significant weight loss or gain” and “item 8: psychomotor agitation or retardation”). These significant DIF values between males and females might have resulted from the natural difference concerning gender on perception or concern about psychological issues and their effects. The results of previous studies of patients undergoing treatment for painful conditions in the emergency department in the United States indicate that female patients presented higher scores for stress and anxiety than male ones (Patel et al., 2014). In addition, the outcomes from a study on the impact of stressful life events on body mass index (BMI) changes also show that stressful life events are associated with an increase in BMI in females only (Udo et al., 2014). The difference in this relationship might be due to DIF across gender.

Four items of the 9Q presented DIF across age groups (“item 1: depressed mood”, “item 2: markedly diminished interest or pleasure”, “item 4: fatigue or loss of energy”, and “item 7: diminished ability to think or concentrate, or indecisiveness”). Previously, Cameron (2013) using other depression measurement tools (the Patient Health Questionnaire (PHQ-9) and the Hospital Anxiety and Depression Scale (HADS)) also found age-related DIF on 3 PHQ-9 items (“little interest or pleasure in doing things”, “feeling down, depressed or hopeless” and “feeling tired or having little energy”), which is consistent with the 9Q items with age-related DIF in our study. This bias between age groups might have resulted from fundamental differences among the adolescent, adults, and elderly groups. Therefore, applying an appropriate tool to measure depression according to differences due to age might have resulted in a reduction in bias in the assessments (e.g., the Children's Depression Inventory (CDI), the Patient Health Questionnaire for Adolescents (PHQ-A), or the Geriatric Depression Scale-15 (GDS-15) (Sheikh and Yesavage, 1986; Trangkasombat and Likanapichitkul, 1997; Johnson et al., 2002).

In addition to gender and age, the items related to a feeling of worthlessness and loss of energy among the elderly presented DIF in the findings from a previous study conducted among the elderly using the GDS-15 across chronic illness groups (Broekman et al., 2008). The DIF across illness was also present in our study for the item related to loss of energy. This might have been because of the impact of different illnesses leading to a difference in fatigue level across participants with and without underlying illnesses.

Several items with DIF attained a high discrimination parameter value to the actual depression trait. Although there are several measurement tools for depression for different settings, ignoring

the difference in the discrimination parameter value of an item in a measurement tool could cause bias. Scoring of the discrimination and threshold parameters across characteristics (e.g., gender, underlying disease, etc.) based on the IRT approach might be useful for reducing bias in depression measurement.

A recent study for the DIF of the PHQ-9 among health care workers found that DIF was not found in any items across age, gender, education and alcohol consumption. They suggested that it might be related with the no to low level of depression for the health care workers (Jiraniramai et al., 2021). Detection of DIF across occupation group is interesting. In addition, recent study on the impact of somatic symptoms on PHQ-9 scores found that several items showed DIF with respect to disease-specific severity, however, the salient DIF was present in very few patients (Katzan et al., 2021). Considering for the impact of DIF related to characteristics could be useful in further study.

The main limitation of this study was the lower number of participants across other interesting variables such as nationality, ethnicity, educational background, or occupation. A further study with a larger sample size should be conducted to determine DIF in other variables and confirm the findings presented in the present study. Moreover, other approaches toward determining the DIF for polytomous items should be considered.

4. Conclusion

In this cross-sectional study to determine the presence of DIF in the responses to the 9Q tool for depression severity assessment among the northern Thai population, DIF was found in the responses for several items according to the participants' characteristics including gender, age, and underlying disease except item 6 and 9. The findings from our study suggest that the IRT approach should be used to determine DIF for polytomous items. In addition, accounting for the difference between the characteristics of participants might reduce the bias in the scoring or assessment of depression severity.

Acknowledgment

We would like to thank the physicians, nurses, medical staffs, and all participants who involved in this study. A primary study on validity of 9Q among northern Thai population was funded by a grant from Mental Department of Mental Health, Ministry of Public Health. This study was partially supported from Chiang Mai University, and Department of Statistics, Faculty of Science, Chiang Mai University.

Ethical approval

This study using de-identified data from the primary study approved by the Ethical Committee, Phra Si Maha Phot Psychiatric Hospital, Ubon Ratchathani, Thailand.

References

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (DSM-5®): American Psychiatric Publishing.

Broekman, B.F., Nyunt, S.Z., Niti, M., Jin, A.Z., Ko, S.M., Kumar, R., et al. (2008). Differential item functioning of the Geriatric Depression Scale in an Asian population. *J Affect Disord*, 108(3):285–290.

Cameron, I. M., Crawford, J. R., Lawton, K., and Reid, I. C. (2013). Differential item functioning of the HADS and PHQ-9: An investigation of age, gender and educational background in a clinical UK primary care sample. *Journal of Affective Disorder*, 147:262–268.

Crane, P. K., van Belle, G., and Larson, E. B. (2004). Test bias in a cognitive test: differential item functioning in the CASI. *Stat Med*, 23:241–256.

Crane, P. K., Gibbons, L. E., Jolley, L., and van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques. DIFdetect and difwithpar. *Med Care*, 44(11 Suppl 3):S115–S123.

Edelen, M. O., and Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*, 16(Suppl 1):5–18.

Gorter, R., Fox, J. P., and Twisk, J. W. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. *BMC Med Res Methodol*, 15:55.

Jiraniramai, S., Wongpakaran, T., Angkurawaranon, C., Jiraporncharoen, W., and Wongpakaran, N. (2021). Construct Validity and Differential Item Functioning of the PHQ-9 Among Health Care Workers: Rasch Analysis Approach. *Neuropsychiatric disease and treatment*, 17:1035–1045.

Johnson, J. G., Harris, E. S., Spitzer, R. L., and Williams, J. B. (2002). The patient health questionnaire for adolescents. *Journal of Adolescent Health*, 30, 196–204.

Katzan, I.L., Lapin, B., Griffith, S., Jehi, L., Fernandez, H., Pioro, E., et al. (2021). Somatic symptoms have negligible impact on Patient Health Questionnaire-9 depression scale scores in neurological patients. *Eur J Neurol*, 28:1812–1819.

Kongsuk, T., Arunpongpaisal, S., Janthong, S., Prukkanone, B., Sukhawaha, S., and Leejongpermpoon, J. (2018). Criterion-Related Validity of the 9 Questions Depression Rating Scale revised for Thai Central Dialect. *Journal of the Psychiatric Association of Thailand*, 63:321–334.

Muraki E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, 16:159–176.

Nering, M.L., and Ostini, R., eds. (2010). Handbook of Polytomous Item Response Theory Models. New York: Routledge Academic.

Patel, R., Biros, M.H., Moore, J., and Miner, J.R. (2014). Gender differences in patient-described pain, stress, and anxiety among patients undergoing treatment for painful conditions in the emergency department. *Acad Emerg Med*, 21:1478–1484.

Raykov, T. , and Marcoulides, G. A. (2018). A course in item response theory and modeling with Stata. College Station, TX: Stata Press College Station.

Reise, S. P., and Haviland, M. G. (2005). Item response theory and the measurement of clinical change. *J Pers Assess*, 84:228–238.

Ritchie, H., and Roser, M., (2018). Mental Health: Our World in Data. Available from: https://ourworldindata.org/mental-health. Cited 10 Mar 2019.

Saracino, R. M., Aytürk, E., Cham, H., Rosenfeld, B., Feuerstahler, L. M., and Nelson, C. J. (2020). Are we accurately evaluating depression in patients with cancer? *Psychol Assess*, 32:98–107.

Sheikh, J. I., and Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist*, 5:165–173.

Snitz, B. E., Yu, L., and Crane, P. K., (2012). Chang CC, Hughes TF, Ganguli M. Subjective cognitive complaints of older adults at the population level: an item response theory analysis. *Alzheimer Dis Assoc Disord*, 26:344–351.

Trangkasombat, U., and Likanapichitkul, D. (1997). The Children's Depression Inventory as a screen for depression in Thai children. *J Med Assoc Thai*, 80:491–499.

Udo, T., Grilo, C. M., and McKee, S. A. (2014). Gender differences in the impact of stressful life events on changes in body mass index. *Preventive Medicine*, 69:49–53

2021 ICMS

INTERNATIONAL CONFERENCE ON COMPUTING, MATHEMATICS AND STATISTICS

e ISBN 978-967-2948-12-4

9 7 8 9 6 7 2 9 4 8 1 2 4