27th ICSEI Congress

Proceedings

Redefining Education, Learning and Teaching in the 21st Century: the Past, Present and Future of Sustainable School Effectiveness

ISBN: 978-979-26-1980-5

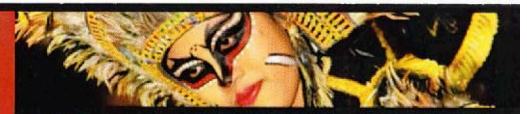
Organized by:

ICSEI International Congress for School Effectiveness and Improvement

Yogyakarta State University

27th ICSEI Congress Proceedings 2014

ICSEI International Congress for School Effectiveness and Improvement



27th ICSEI Congress, Yogyakarta, Indonesia 2-7 Jan 20

27th ICSEI Congress

Proceedings

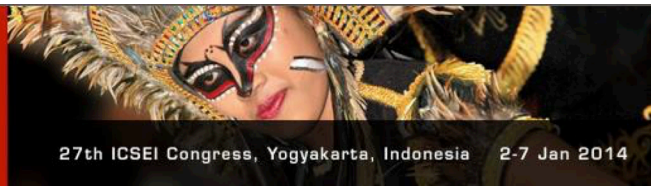
Redefining Education, Learning and Teaching in the 21st Century: the Past, Present and Future of Sustainable School Effectiveness

Organized by:

ICSEI International Congress for School Effectiveness and Improvement

Yogyakarta State University

ICSEI
International Congress for School Effectiveness and Improvement



ITEM RESPONSE THEORY (IRT) AS A MEASUREMENT MODEL TO ASSESS STUDENTS' CHARACTERISTICS IN BROAD AREA: A STUDY ON DICHOTOMOUS AND POLYTOMOUS MULTIPLE-INTELLIGENCE INSTRUMENT

Farida Agus Setiawati
Yogyakarta State University
faridaagus@yahoo.co.id

Abstract

Measurement model is the procedures to analyze the characteristic of instrument. There are classical and modern measurement models. Classical theory model has been used so far in assessing students' characteristics, but it has a weakness in the analysis of the characteristic instrument, because it is group estimation or bound to the sample. The weakness of this model is solved by the modern theory called Item Response Theory (IRT). IRT has an advantage, in which it is parameter estimation so that it can be used in a population or in a broad area. This study aims to describe a model of IRT measurement on multiple intelligences (MI) instrument that can be used in a broader area.

The MI instrument was designed and analyzed into two models, dichotomous and polytomous IRT models. The item analysis was done using graded response model (GRM) with parscale program to examine the fit model of items and estimate the parameters of index difficulty and item discrimination. The results of parameter estimation were used to estimate the parameter of ability (θ) and item-information function.

The findings suggest there is fit model of some items. The model of the fit items can be used to measure the characteristics of student' intelligences in a broad area. Futher, index of difficulty and item dicrimination of each item were analyzed. The result of the analysis show that the index difficulty of dichotomous items is higher than polytomous items, the power discrimination of dichotomous is less than polytomous items, the items information function of dichotomous items is less than polytomous items. Thus, the measurement using polytomous items are more accurate than dichotomous items.

A. Introduction

Measurement is giving score or number to the learning outcomes and student characteristic. There are two models of measurement theory, the classical and modern theory. The assumption of classical theory is about true score. The observed score is sum of true score and error score measurement. If we measure in several times, the observeb score is same as the true score, and the error score is zero. Thus, the observed score can be used to interpret the ability of student or examinee.

The classical theory is simple in administration and analysis characteristic of instrument. This model is often used in the assessment of learning outcomes and the identification of students' abilities. Although this model is used in many assesments, it has a weakness. The item parameter depends on person or population, and the parameter person depends on item (Hambleton, Swaminathan, & Rogers, 1991: 2-5). In classical theory, the item difficulty of instrument was high if it is done by the low ability of group sample and it was low if it is done by the hight ability of sample. Pollard, Dixon, Dieppe dan Johnston, (2009: 3) also state that IRT has more advantages than classical theory because of used invarian sample, so the item parameter does not depend on level ability sample. This weakness was solved by items response theory (IRT). This theory is parameter estimation. The instrument characteristics are not dependent on the characteristics of groups or samples, but it can be used in population. IRT model estimate every person and item. Thus, the modern theory approach is considered more powerful than the modern theory.

The social insruments, especially psychological and educational science, were developed in dychotomous and polytomous format, therefore, the multiple intelligence instrument is used in this research. MI instrument has been was developed by several researchers from many countries, for example, Phyllis Reardon and Isabell Dyke from Canada, Thomas Amstrong from Virginia, Charles Branton Shearer from USA. The scoring and analyzing of the MI instrument use classical theory, therefore it must be tried out if we want to use it in other samplse and person in other places. The modern theory or IRT can be used to generelize in the population or other plece. The purpose of this study was to develop an instrument of MI and analyze it using IRT, so that it can be used by students in various places or broader area

Item response theory (IRT) was popularized by Hambleton and Swaminathan in 1985. IRT models use mathematical concept which the probability correctly answering an item depends on the ability of the subject and item characteristics. In this case, a person who has the ability or latent trait will provide higher response at a different point than someone who has a low capacity. Related to

item characteristics, IRT analysis is invariant, meaning i parameter does not depend on how well the sample, and vice versa. Thus, subject scores will not change despite being in the group with high or low ability.

The measurement model of IRT can be grouped into two models, dichotomous and polytomous model. Dichotomous model is a model that uses two responses, like right and wrong. The score is one (1) if the examinee answesr correctly, and zero (0) if the examinee fails or wrong in answer. Score of 1 and 0 can also show the response of examinee, that score 1 indicates higher response than 0. There are many kinds of instruments that use traditional dichotomous models, including: right-wrong, agree-disagree, appropriate-inappropriate, and yes-no.

Dichotomous item response theory is developed by using three logistic models, i.e. one parameter, two parameters, and the three parameters logistic model. One parameter logistic model uses only one parameter, which is item difficulty index. Two parameter logistic model uses parameters of item difficulty index and power of differences. Three parameter logistic models uses item difficulty index, power of discrimination, and guessing or pseudoguessing.

There are several IRT models used in scoring polytomous, which are graded response model (GRM), modification of response model or modified graded response model (M-GRM), partial credit model (PCM), generalized partial credit model (G-PCM), rating scale model (RSM) and nominal response models (NRM) (Embretson & Reise, 2000:95). M-GRM is based on 2 parameter logistic model. G-PCM and PCM use Rasch models or 1 parameter, RSM uses location-scale model of Andrich and NRM uses for irregular response categories.

Polytomous of IRT models that can be used to measure the Likert scale are graded response model (GRM). This model was developed by Samejima 1969 (Ostini, 2006: 61, van der Linden & Hambleton, 1997: 86). Graded response model is used when the participant responds to item category scores and the scores tended to increase sequentially.

B. Methods

The study develops the MI instrument using the following steps: determining the construct and specification of the instrument, writing items, trying out the instrument, and analyzing the items using IRT. The item analysis was done to estimate the parameters of index difficulty, item discrimination, and item fit model. The estimation of parameters was analyzed using the graded response model (GRM) with parscale program. The result of parameter estimation were used to estimate the parameter

of ability (θ) and item-information function. The source of research data is the data of multiple intelligences instrument on 443 student.

C. Results

Fit model of items was tested to Yogyakarta State University Students. There were 16 of 141 unfit items and they were dropped from the instrument. Several fit items also dropped because there should be only 72 items used in the instrument. Further, 72 items were analyzed again to calculate the parameters items.

The IRT was analyzed using pascsale model to estimate index of difficulty, discrimination power and information function. There were 4 kinds of parameter b between categories of polytomous items that moved from very easy to difficult. Parameter b1 on all items had scores below -2, b2 mostly ranging from -2 to 0, b3 was ranging from 0 to 2, and b4 above 2. Overall, the average index difficulty of items was -0.956. The average category b1 = -3.481, b2 = -1.042, b3 = 1.088, and b4 = 2.902. In general, this instrument had a moderate level of item difficulty. Parameter of power discrimination was all items above 0.2. It can be said that the items of instrument could distinguish people who had high and low ability.

The dichotomous items were analyzed using parscale program. The average of parameters b1 was -0,149 and the average of power discriminations was 0,533. The index difficulty of dichotoumous items was higher than polytomous items, and the power of discrimination dichotomous was less than polytoumous items.

Item information function (IIF) and test information function (TIF) was the measurement term in IRT to indicate the accruration of item or test. Tabel 1. was score of TIF dichotomous dan polytomous format instrument. It showed that TIF of polytomous format was higher than dichotomous format and the standart error measurement of polytomous format was smoller than dichotomous format. Therefore, the polytomous items were more accurate than dichotomous items.

Table 1. Test Information Fuction and Standar Error Measurement of Dimenton instrument

No	Dimention	Dichotomous			s
		TIF	SEM	TIF	SEM
1.	Linguistic	1.177	0.922	7.048	0.377
2.	Logic-mathermatic	2.214	0.672	9.259	0.329
3.	Visual-spatial	2.869	0.590	10.624	0.307

4.	Music	3.987	0.501	12.183	0.286
5.	Kinestetic	4.448	0.474	12.718	0.280
6.	Intrapersonal	3.519	0.533	10.780	0.305
7.	Interpersonal	2.935	0.584	9.747	0.320
8.	Naturalistic	1.744	0.757	7.836	0.357
9.	Existencial	1.698	0.767	8.216	0.349
Average		2,732	0,644	9,823 0.3	23

D. Discussion

The fit model in IRT analysis indicates that the item can predict parameters instrument. The parameters of fit item can be used in the large area. IRT analysis is beneficial in the study of the instrument that is used in a wide area with the large number of subjects. It is related to the invariant nature of the IRT analyzes where the resulting score was not associated with person characteristics. The results of IRT analysis are not required to be tried again to other persons in other places. Thus, standardized instruments used in broader area such as the UN tests, TPA tests, and psychological tests, will be appropriate if analyzed using IRT.

Index discrimination of dichotomous items of multiple intelligences instrument is smaller than polytomous items, but index of difficulty of dichotomous items is higher than polytomous items. The differences in parameters are the format of dichotomous items using Thurstone type and the format of polytomous items using Likert type. The Likert type has high in social desirability. The answer, thus, may not appropriate with the real answer or the examinee can do faking in answer. Examinee may choose the highest respons eventhought the real performance does not reflect it. This condition causes the high index difficulty of Likert type. The number of response also influences the index of discrimination and informationn function items. The more response of an item, the more index discrimination and item information function. Grassi, et. al (2007, 478), states information items in Likert type decreases if the respose changes from five response to binary format. Eko Hariadi (2007; 91-92) also points out the average of information function in polytomous response is higher than dichotoous response.

Acnowledgments

Prof. Djemari Mardapi and Prof Saifuddin Azwar as Promotors

Referencess

Ekohariadi. (2007). Penskalaan butir format respons pilihan dan respons bebas berdasarkan model RASCH dan partial kredit. *Disertasi*, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta

Embretson, S. E., & Reise, S. P., (2000). *Item response theory for psychology*. New Jersey: Lawrence Elbaum Associates, Publisher.

Grassi, M., Nucera, A., Zanolin, E., Omenaas, E., Anto, J. M., & Leynaert, B. (2007). Performance comparison of Likert and binary formats of SF-36 version 1.6 Across ECRHS II adults populations. *Value in Health,* 10,6, 478-488.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory : Principles and Aplication*. Boston: Kluwer.

Hambleton, R. K., & Zaal, J. N. (1991). *Advance in educational and psychological testing*. Boston: Kluwer Academic Publisher Group.

Likert, R. (1932). *Technique for the Measurement of Attitudes. Archives of Psychology*. 22,140-146.

Ostini, R., & Nering, M. L., (2006). *Polytomous item reponse theory models*. London: Sage Publication.

Pollard, B., Dixon, D., Dieppe, P., & Johnston, M. (2010). Measuring the ICF components of impairment, activity limitation and participation restriction: an item analysis using classical test theory and item response theory. *Health and Quality of Life Outcomes*,7,41.

Thurstone, L. L. (1927). Method of paired comparison for social values. *Journal of Abnormal and Social Psychology*, 21, 384-400.

Van der Linden, W.J., & Hambleton, R.K., (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag New York Inc.