**International Journal of Evaluation and Research in Education (IJERE)**
Vol.7, No.4, December 2018, pp. 313~316
ISSN: 2252-8822 DOI: 10.11591/ijere.v7i4.15146 

# Using Partial Credit Model to Improve the Quality of an Instrument

**Saras Krishnan[1], Noraini Idris[2]**
[1] INTI International University, Malaysia
[2] Universiti Malaya, Malaysia

**Article Info**

***Article history:***

Received Aug 21, 2018
Revised Sep 4, 2018
Accepted Oct 10, 2018

***Keyword:***

Construct validity
Content validity
Fit statistics
Reliability
Separation

**ABSTRACT**

In using the Rasch model to improve the quality of an instrument, analysis purports to determine if the sample collaborates well with the items in the instrument such that the results are measuring a single underlying variable. The relevant properties of Rasch analysis are reliability and validity which are key indicators of the quality of a measurement instrument. This paper discusses the use of one type of Rasch model that is the Partial Credit Model to investigate reliability and validity of an instrument. By removing or changing items in the instrument when conditions of reliability and validity are not met, the quality of the instrument is improved.



***Corresponding Author:***

Saras Krishnan,
Faculty of Engineering and Quantity Surveying,
INTI International University,
Persiaran Perdana BBN, Putra Nilai, 71800 Nilai, Negeri Sembilan, Malaysia.
Email: saras.krishnan@newinti.edu.my

## 1. INTRODUCTION

Assessment is an important process in any curriculum primarily because assessment drives pedagogy. It is thus highly crucial to develop instruments that are able to accurately represent the achieved learning outcomes. Despite the need to conduct reliable and consistent measurements, researchers had pointed out the lack of high quality instruments to assess achieved learning outcomes [1]. The major concern in the development of an instrument is if it actually measures what the researcher sets out to measure [2]. The two important measurements of the quality of an instrument are the reliability and validity of the instrument. This paper discusses the process of using Rasch model, in particular the Rasch Partial Credit Model to improve the quality of an instrument. Prior to discussing the reliability and validity criteria of the Partial Credit Model (PCM) in the next section, in this section we will first have a look at the general conditions for reliability and validity of Rasch models.

An instrument is reliable when it maintains consistency in different settings and the measurement results are generalizable to other samples [2, 3]. Using Rasch models, the measure of item reliability is in the range of 0.00 to 1.00 whereby the higher the value, the higher the reliability is [2]. A point to note is that reliability does not neccesarily imply validity. However, unreliability detracts from validity [3] and so reliability is an important measure of the quality of an instrument. Another measure of reliability is the Cronbach's alpha value whereby values more than 0.70 are deemed acceptable for aggregate data [4]. However, it has been argued that exceptionally high Cronbach's alpha value indicates redundancy and thus ideally the value should be between 0.70 and 0.90 [3]. Conceptually, item and case reliabilities are equivalent to Cronbach's alpha [5].

The two types of validity that can be measured using the Rasch models are the construct validity and the content validity. Construct validity is used to determine if an instrument measures the unobservable construct that it purports to measure [3], and if the items in the instrument are systematically and predictably related to each other along the construct that is being measured [6]. This is achieved by examining the relationship between the measure being evaluated and the variables known to be related or theoretically related to the construct being measured by the instrument [2]. On the other hand, content validity determines if the instrument measures the entire domain related to the construct it was designed to measure [7].

## 2. PARTIAL CREDIT MODEL (PCM)

Of the different types of Rasch models, the Partial Credit Model (PCM) is particularly useful for instruments with polytomous scoring because PCM allows varying levels of intermediate correctness between completely correct and completely wrong answers. In other words, different items in the instrument can have different number of response options. As such, PCM is an ideal measurement tool in the educational contexts because it is common to award part marks to students' partly correct answers whereby higher marks represents higher learner ability. Furthermore, from the measurement perspective polytomous data is more desirable than dichotomous data because the error estimates are smaller for polytomous data for the same number of items.

In using PCM to determine the quality of an instrument, we investigate the values of the fit statistics [8], in particular the mean square fit statistics and the standardized fit statistics. Infit mean square is used to investigate the size of the randomness or the amount of distortion of a measurement system. Infit mean square more than 1.20 suggests unusual or inappropriate response patterns while infit mean square less than 1.00 indicates too little variation in the response patterns that further suggests presence of redundant items [5]. The standard deviation of less than 2.00 indicates an overall fit of the items [5].

Further, the other two properties used to investigate the quality of a measurement instrument are the item separation and the mean measure. Item separation measures the spread of items in standard error units [9]. It reflects the extent to which the items are separated along a scale in terms of their difficulty levels. An item separation of more than 1.00 indicates that there is enough spread of items and thus the instrument is acceptable. Mean measure is an indicator of the difficulty level of the instrument whereby a positive value suggests that the instrument is easy while a negative value suggests that the instrument is difficult. In specific, mean measure of values 1.00 and 2.00 indicate that the instrument is too easy whereas mean measures -1.00 and -2.00 indicate that the instrument is too difficult. In both cases, it is advisable to have the instrument revised [5].

The construct validity of an instrument is determined by inspecting the table of misfit order of items. Construct validity ensures that a single construct is being measured, and that the items are systematically and predictably related to each other [6]. The conditions for construct validity are: (1) infit mean square between 0.77 and 1.30 [6, 10], and (2) infit z-standardized between -2.00 to +2.00 [9]. These conditions were adopted in earlier papers as well (see, [11, 12]). The content validity is determined using the point measure correlation values whereby these values should be positive.

## 3. AN EXEMPLAR

In this section, we discuss an example of an instrument that was refined using the PCM measurements (see, [13]). Discussion is supported by the tables of summary of statistics and misfit order of items which have been generated using the Winsteps companion to Bond and Fox (2007) [9].

### 3.1. Measurement of Reliability

Table 1 displays the reliability values, the separation values, the infit mean square, the standard deviation, the mean measure and Cronbach's alpha value for the three successive stages of development of an instrument.

Table1. Summary of statistics

| | Reliability | Separation | Infit MNSQ | S.D. | Mean | Cronbach |
|---|---|---|---|---|---|---|
| Stage I | 0.97 | 5.48 | 1.02 | 0.23 | 0.15 | 0.85 |
| Stage II | 0.98 | 7.22 | 1.00 | 0.09 | -0.22 | 0.75 |
| Stage III | 0.98 | 6.48 | 1.00 | 0.08 | -0.18 | 0.75 |

The item reliabilities for the three stages of the instrument development are very high, that is more than 0.90, indicating high reliability. This is supported by Cronbach's alpha values which are more than 0.70. Also, Cronbach's alpha values between 0.70 and 0.90 indicate non-redundancy of items in the instrument [3]. In addition, the item separation values of more than 1.00 indicate enough spread of items. With reference to Fisher's (2007) instrument quality criteria, the item reliability values of more than 0.94 together with the item separation values of more than 5.00 indicate that the instrument has excellent quality [14], as is the case in all three stages shown in Table 1. The infit mean square are between 1.00 and 1.20 indicating that the response pattern is suitable, there is enough variation in the responses and there are no redundant items in the instrument. The standard deviation values for the items are less than 2.00, an indication of an overall fit of the items [5]. Further, the items are not too easy or too difficult, as indicated by the values of the mean measure, and therefore the instrument is acceptable for measurement purposes.

### 3.2. Measurement of Validity

Table 2 displays the infit mean square (MNSQ), the infit z-standardized (ZSTD) and the point measure correlation (PMC) values which were used to establish the validity of the instruments.

Table 2. Misfit order of items

| | MNSQ | ZSTD | PMC |
|---|---|---|---|
| Stage I | 0.80 < MNSQ < 1.92 | -1.10 < ZSTD < 3.50 | 0.27 < PMC < 0.70 |
| Stage II | 0.82 < MNSQ < 1.19 | -1.60 < ZSTD < 1.50 | -0.19 < PMC < 0.57 |
| Stage III | 0.86 < MNSQ < 1.17 | -1.60 < ZSTD < 1.40 | 0.04 < PMC < 0.52 |

The instrument in Stage I did not meet the conditions of construct validity since some of the infit mean squares are more than 1.30. By removing or modifying some items, the instrument in Stage II fulfilled the conditions of construct validity whereby 0.77 < MNSQ < 1.30 and -2.00 < ZSTD < 2.00. However, the instrument in Stage II did not fulfill the condition for content validity because there are negative PMC values. Further refinement of the items led to the instrument in Stage III which have better quality than the instruments in Stage I and Stage II in terms of reliability and validity based on the criteria discussed in this paper.

## 4. DISCUSSION

The use of fit statistics is influenced by its intended application whereby for a high stakes testing, the reasonable stipulated range of the fit statistics are between 0.80 and 1.20, for a survey the values are between 0.60 and 1.40 and for clinical observation the recommended values of the fit statistics are between 0.50 and 1.70 [15]. Linacre (2002), in discussing the implications of using different values of the mean square fit statistics and standardized fit statistics stated that the mean square fit statistics between 0.50 and 1.50 is productive for measurement whereas standardized fit statistics between -1.90 and 1.90 indicates that data have reasonable predictability [16].

Over the years, different sets of criteria of reliability and validity have been adopted. For instance, Aoyama (2007) set the condition of content validity to be 0.77 < MNSQ < 1.30 [6] whereas Pada, Kartowagiran and Subali (2016) used the condition between 0.70 < MNSQ < 1.30 [17]. On the other hand, Zubairi and Kassim (2016) used the condition 0.40 < MNSQ < 1.60 [18] for acceptable fit. Also, different researchers have reported different measures in their discussion. For instance, Watson and Callingham (2003) [10] did not examine the outfit mean square statistics in their study but Sabudin, Mansor, Meerah and Muhammad (2018) [19] did.

## 5. CONCLUSION

Rasch model is a valuable tool to investigate the quality of an assessment instrument in terms of its reliability and validity. Instead of having to completely revise the instrument, Rasch model informs us which items in the instrument specifically violates the conditions of reliability and validity. By removing or revising the items in concern, the quality of the assessment instrument is improved so that the instrument fits the purpose of assessment. Researchers have used and reported different measures of Rasch analysis in their studies as seen in the discussion above.

In conclusion, there are no hard and fast rules on the acceptable range of these measures neither a condition of which of these measures are obligatory. Essentially, the researcher makes the ultimate decision on how to use the properties of Rasch analysis to improve the quality of a measurement instrument. However, regardless of the measures and criteria adopted, the Rasch Partial Credit Model which is especially

useful when dealing with polytomous data is an excellent tool to identify possible threats to the reliability and validity of an instrument and thus improving the quality of the instrument.

## REFERENCES


[1] Garfield J, Ben-Zvi D. "How students learn statistics revisited: A current review of research on teaching and learning statistics". *International Statistical Review*. 75(3):372-96; 2007.
[2] Hagan TL. "Measurements in quantitative research: How to select and report on research instruments". *Oncology Nursing Forum*. 41(4):431-433; 2014.
[3] Pesudovs K, Burr JM, Harley C, Elliott DB. The development, assessment, and selection of questionnaires. *Optometry and Vision Science*. 84(8):663-74; 2007.
[4] Ismail W, Zailani MA, Awad ZA, Hussin Z, Faisal M, Saad R. Academic and Social Media Practices of Arabic Language among Malaysian Students. *Turkish Online Journal of Educational Technology (TOJET)*. 16(3):1-0; 2017.
[5] Green KE, Frantom CG. "Survey development and validation with the Rasch model". In *International Conference on Questionnaire Development, Evaluation, and Testing*, Charleston, SC, 2002 Nov 14.
[6] Aoyama K. "Investigating a hierarchy of students' interpretations of graphs". *International Electronic Journal of Mathematics Education.* 12;2(3):298-318; 2007.
[7] Heale R, Twycross A. "Validity and reliability in quantitative studies". *Evidence-based Nursing*: ebnurs-2015.
[8] Callingham R, Carmichael C, Watson JM. "Explaining student achievement: The influence of teachers' pedagogical content knowledge in statistics". *International Journal of Science and Mathematics Education*. 14(7):1339-1357; 2016.
[9] Bond TG, Fox CM. *Applying the Rasch model: Fundamental Measurement in the Human Sciences*. 2nd edition. New Jersey: Lawrence Erlbaum; 2007.
[10] Watson J, Callingham R. "Statistical literacy: A complex hierarchical construct". *Statistics Education Research Journal*. 2(2):3-46; 2003.
[11] Krishnan S, Idris N. "The Use of a Hierarchical Construct to Investigate Students' Learning of Inferential Statistics". In *Proceedings of the Joint IASE/IAOS Satellite Conference*, Macao, China 2013 Aug.
[12] Krishnan S, Idris N. "Investigating reliability and validity for the construct of inferential statistics". *International Journal of Learning, Teaching and Educational Research*. 4(1):51-60; 2014.
[13] Krishnan S, Idris N. "Improving the Quality of a Measurement Instrument using the Partial Credit Model". In *International Conference on Learning and Teaching*, Singapore, 2015 March 25 & 26.
[14] Fisher WP. "Rating scale instrument quality criteria". *Rasch Measurement Transactions*. 21 (1), 1095; 2007.
[15] Wright B. "Reasonable mean-square fit values". *Rasch Measurement Transactions*. 8:370; 1994.
[16] Linacre JM. "What do infit and outfit, mean-square and standardized mean". *Rasch Measurement Transactions*. 16(2):878; 2002.
[17] Pada AU, Kartowagiran B, Subali B. "Separation index and fit items of creative thinking skills assessment". *Research and Evaluation in Education (REiD)*. 1;2(1):1-2; 2016.
[18] Zubairi AM, Kassim NL. "Classical and Rasch analyses of dichotomously scored reading comprehension test items". *Malaysian Journal of ELT Research*. 7;2(1):20; 2016.
[19] Sabudin S, Mansor AN, Meerah SM, Muhammad A. "Validity and Reliability of Students' Science and Technology Culture Instrument (BST-M) using Rasch Measurement Model". *International Journal of Academic Research in Business and Social Sciences*. 8(5):986-995; 2018.


## BIOGRAPHIES OF AUTHORS


Saras Krishnan obtained her doctorate degree in Mathematics Education from the University of Malaya, Malaysia in 2014. She has a BSc in Industrial Science majoring in Mathematics and MSc in Applied Mathematics, both from the Universiti Teknologi Malaysia, Malaysia. She is a full time teaching staff at the INTI International University in Malaysia and over the years have taught different courses including calculus, engineering mathematics and mathematical studies for various programs. Her research interests include teaching and learning, and assessments, particularly in the higher learning institutions.
E-mail: saras.krishnan@newinti.edu.my



Prof Dato' Dr Noraini Idris holds a BSc Ed. (Hons) majoring in Mathematics and Physics and M.Ed. in Mathematics Education from the Universiti Malaya, Malaysia. She went on to obtain a PhD in Mathematics Education from the Ohio State University in the United States of America. She was awarded the Distinguished Diversity Enhancement Award when she was in USA for organizing workshops to improve performance of the minority groups. She is also a gold medal recipient at the International Innovation and Invention Competition held at Geneva, Switzerland in 2005.
E-mail: n.idris07@gmail.com