Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM

Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

JOURNAL OF INFORMATION SYSTEM AND TECHNOLOGY MANAGEMENT (JISTM)
www.jistm.com

GLOBAL ACADEMIC EXCELLENCE

RASCH VALIDATION OF THE WBISA APPLICATION QUESTIONNAIRE IN HIGHER LEARNING INSTITUTION

Nur Razia Mohd Suradi[1*], Azlinda Abdul Aziz[2*], Saliyah Kahar[3], Nor Azliana Akmal Jamaludin[4], Rahayu Handan[5], Nik Nordiana NAR[6]

1 Faculty Communication, Visual Art and Computing, Universiti Selangor, Malaysia
Email: razia@unisel.edu.my
2 Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Melaka Campus Jasin, 77300 Merlimau, Melaka, Malaysia
Email: azlindaaziz@uitm.edu.my
3 School of Computer Science, Faculty of Innovation and Technology, Taylor's University, Malaysia
Email: saliyah.kahar@taylors.edu.my
4 National Defence University of Malaysia (UPNM), Malaysia
Email: azliana@upnm.edu.my
5 Faculty Communication, Visual Art and Computing, Universiti Selangor, Malaysia
Email: raha@unisel.edu.my
6 Faculty Communication, Visual Art and Computing, Universiti Selangor, Malaysia
Email: niknordiana@unisel.edu.my
* Corresponding Author

Article Info:

Article history:
Received date: 05.01.2025
Revised date: 19.01.2025
Accepted date: 10.02.2025
Published date: 03.03.2025

To cite this document:

Suradi, N. R. M., Abdul Aziz, A., Kahar, S., Jamaludin, N. A. A., Handan, R., & Nar, N. N. (2025). Rasch Validation Of The Wbisa Application Questionnaire In Higher Learning Institution. *Journal of Information System and Technology Management, 10* (38), 01-11.

Abstract:

It is important to access the quality of any academic application in higher learning institution. Basically, users of this application are academicians who rely on system to perform daily activities. However, limited study has been conducted to investigate this matter. Constructing an item questionnaire with a content validity, reliability and fairness is not easy. The questionnaire was then distributed among the academicians who used WBISA application via online platform. Therefore, this study aims to reveal to develop the item of the questionnaire using Rasch measurement model. The questionnaire consists of 61 items used with 9 categories of characteristics. Overall, the results have shown that item fits were within an acceptable range. Meanwhile, the Cronbach's alpha was a good value at 0.97, with person reliability of 0.96 and item reliability of 0.87. The person separation was observed at a value of 5.05 while the item separation is considered acceptable at a value of 2.54; indicating good measurement capability which supports its use in evaluating academician's using WBISA application. Finally, a well-developed of item questionnaires has been established.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

DOI: 10.35631/JISTM.1038001

Keywords:

Academic Application; Quality Model; Rasch

Introduction

In the era of digital education in higher learning, most of the applications have been developed in-house. Each of these applications has their own purpose. One example of the academic application is WBISA application. WBISA application is a web-based application developed using PHP; however, there is also other approach which could be used, such as machine learning (Nazurah et al., 2023). The construction of item questionnaire is vital to evaluate the academic application. However, constructing an item questionnaire is a challenging process because each of the item questionnaires has their own purpose and is designed to measure specific aspects. A good assessment must involve a good instrument (Natsir et al., 2022).

A Rasch model is unidimensional and adopts that the response to a specific scale item is a function of that item's difficulty and its respondent's ability (Rasch, 1961, 1960). According (Bond & Fox, 2015), Rasch is a measurement model capable in solving the issue of validity by providing useful statistics and able to perform validity. The validity and the reliability of the questionnaire is very crucial in maintaining the accuracy of the questionnaire (Ariffin et al., 2010; Balapumi et al., 2015; Mark Wilson, 2005). With the Rasch model, researchers can assess each item's performance and its relationship to others. Biased or unreliable items are recognized and can be improved or removed. The Rasch model also establishes a common measurement scale, allowing accurate calibration of item difficulty and accurate assessment of individual abilities, leading to fair assessments.

Several studies showed that Rasch can perform analysis greatly. (Natsir et al., 2022) used Rasch in testing the instrument and measuring the ability of students in junior high school mathematics courses, especially in Algebra material. Another researcher applied Rasch in developing quality model for academic application in higher learning (Mohd Suradi, 2022) while (Banawi et al., 2023) applied Rasch in measuring the quality of the items from the Final Semester Examination on the Basic Science Concepts course. Furthermore, a study conducted by (Bilqis et al., 2024)aimed to test the understanding of high school students concerning the dangers of radiation exposure from the nearby environment.

This study will show how to practice Rasch analysis to create an item questionnaire and appraise the validity and reliability of the item questionnaire. By discussing this construction process, the study intends to provide other researchers with a strong understanding of how to create an item questionnaire that is calibrated on common logit scale, facilitating fair and accurate assessments across various proficiency levels.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

Research Methodology

Item Development

This section outlines the approaches employed to produce the questionnaire. This study employed a quantitative approach which involved item questionnaire construction process. The item questionnaire consists of 61 items with 4-Likert scale. The questionnaire survey for this research consisted of two (2) sections: Respondent Profile and list of quality characteristics. The original Likert scale comprised five symmetrical and balanced points. However, throughout the year, the measurement range in relations options vary from two-points to eleven-points (Taherdoost, 2019). The questionnaire used the Likert scale (“1-Strongly Disagree”, “2-Disagree”,”3-Agree”, “4-Strongly Agree”). Figure 1 presented the item questionnaire used in this study.

No.	Characteristics	Strongly Disagree	Disagree	Agree	Strongly Agree
1	**Usability characteristics** Definition: The capability of software to assist user in using the software				
a.	The application is easy to use.	1	2	3	4
b.	The term used in the application easy to understand.	1	2	3	4
c.	The information display is organized accordingly.	1	2	3	4
d.	The information in the application easy to access.	1	2	3	4
e.	The structure of the application is easy to understand.	1	2	3	4
f.	The function provides easy to understand.	1	2	3	4
g.	The navigation gives correct information.	1	2	3	4
h.	The application can be accessed easily.	1	2	3	4

Figure 1: Survey Questionnaire

In Figure 2 below, there are eight steps involved in developing quality characteristics and item questionnaire for WBISA application. The process of identifying the characteristics of web-based educational applications is shown as in Figure 2.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

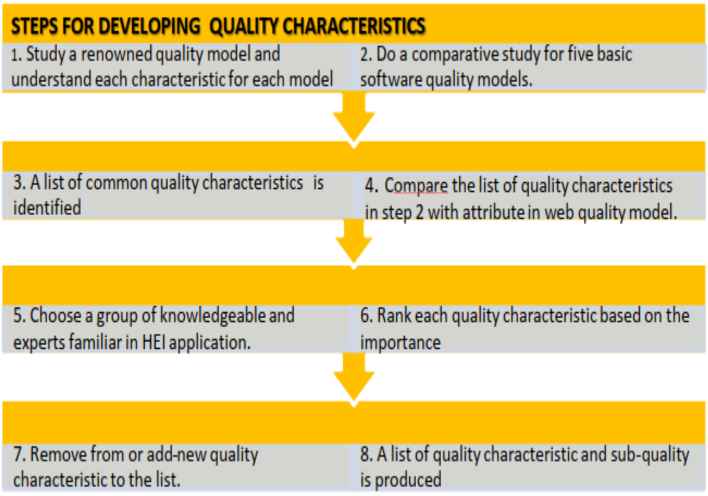

Figure 2: Steps In Developing Quality Characteristics

The first is to study a renowned quality model and understand each characteristic for existing quality model. The second is to do a comparative study for five basic software quality models. Software engineering consists of the McCall quality model, Boehm quality model, FURP quality model, Dromey's quality model and ISO 9126 quality model. The third is to list common quality characteristics by listing all the characteristics and sub-characteristics from the quality model. Fourth is to compare the list of quality characteristics in step 2 with attributes in the web quality model. The new list is produced by combining the two categories of quality models. Fifth is to choose a group of knowledgeable and experts familiar with the HLI application. The sixth is to ask the experts to rank each quality characteristic based on its importance. Seventh, the process of removing from or adding new quality characteristics to the list. Each quality characteristics is attached with the set of item questionnaire with the related definition. Finally, a list of quality characteristics and sub-quality factors is produced. The detail process has been discussed by (Mohd Suradi et al., 2018).

The success of questionnaire development not only depends on the correct grammar and content but also the question to be asked to ensure the respondents are willing to answer the questionnaire correctly (Zikmund et al., 2013). The process in developing the questionnaire involves seven processes as depicted in the Figure 3.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

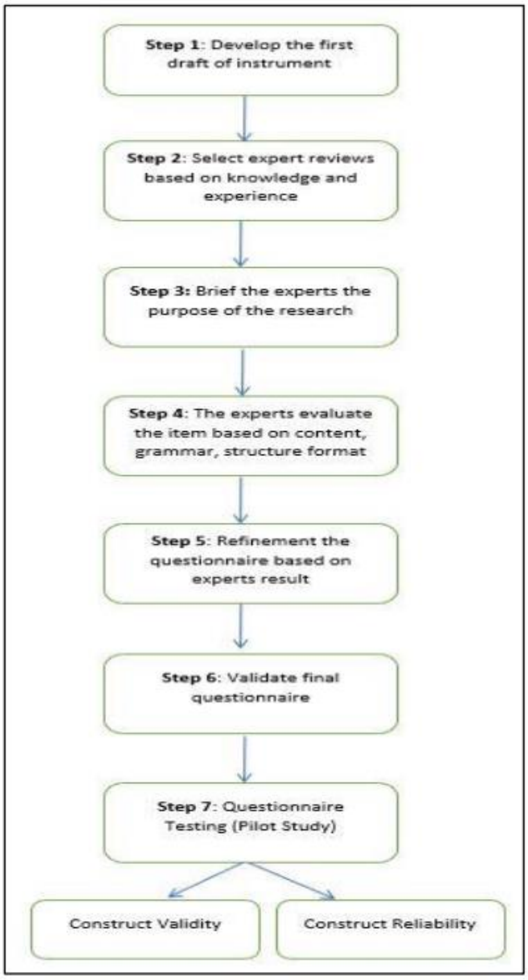

Figure 3: Steps in Questionnaire Development

Participant Selection

The sampling technique is a method where the data is collected from all cases in the population. The participant is selected using clustering sampling based on the category of higher learning institution. (Singh, P. et al., 2015) described that cluster sampling refers to a situation where the selected groups have similar characteristics and are chosen randomly. Participant is from academicians who used WBISA application. To avoid responses that are not biased, the survey questionnaires were distributed in the Public Institution implementing WBISA application. Phone calls, emails, social media platforms such as Facebook and face-to-face interactions were used in distributing the questionnaire. (Taherdoost, 2017) stated that in most social and management surveys, the response rate for e-mailed surveys is hardly 100%.

Results

The test results obtained are in the form of scores that was analyzed using Winsteps Rasch software (Linacre, 2009) which aims to determine fit and misfit items. The results from the 55 respondents (extreme and non-extreme) were tabulated and analysed. The summary statistic report with an extreme person (N=55) showed that only one person gave an extreme response. The extreme person agreed with all the items. The Cronbach's alpha was at 0.97, indicating good internal consistency reliability for familiarity using web-based academic applications and the characteristics of the application. The person reliability did not have a difference between

Journal of Information System and Technology Management
EISSN: 0128-1666
JISTM
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

person reliability before the data was cleaned. The person reliability was 0.96 and the person separation index contributed a fair separation of 5.05 *logit* as can be referred to in Table 1. The person separation index indicates how effectively a set of items can distinguish between individuals being measured. In this study, the results show five distinct levels of person separation in their responses to the items.

According (Banawi et al., 2022; Bilqis et al., 2024), the criteria in determining the value of Item Reliability and Person Reliability in the Rasch model are based on the subsequent criteria as below:

a. > 0.94: EXCELLENT
b. 0.91 – 0.94: VERY GOOD
c. 0.81 – 0.90 : GOOD
d. 0.67 – 0.80 : FAIR
e. < 0.67: POOR

Table 1: Person Summary Statistic Before and After Data was Cleaned

	Before data was cleaned (n=61)	**After data was cleaned (n=54)**
Cronbach Alpha(α)	0.96 (Good)	0.97 (Good)
Person		
Reliability Index	0.95	0.96
Separation Index	4.58 (Very Good)	5.05 (Very Good)
Mean	0.76	1.03
Standard Deviation (SD)	0.03	1.44
Infit MNSQ	1.00 (Very Good)	1.00 (Very Good)
Infit ZSTD	-.5	-.3
Max	4.38	4.96
Min	-1.12	-1.49
Model error	0.21 (Excellent)	0.25 (Very good)

Table 2: Item Summary Statistic Before and After Data was Cleaned

	Before data was cleaned (n=61)	**After data was cleaned (n=49)**
Cronbach Alpha(α)	0.96	0.97
Item		
Reliability Index	0.88 (Good)	0.87 (Good)
Separation Index	2.66 (Fair)	2.54 (Fair)
Mean	0.00	0.00
Standard Deviation (SD)	0.60	0.71
Infit MNSQ	0.98 (Very Good)	0.98 (Very Good)
Infit ZSTD	-.1	-.1
Max	1.26	1.73
Min	-1.13	-1.36
Model error	0.20 (Very Good)	0.25 (Very Good)

The values of INFIT and OUTFIT mean squares of the response category are reported as shown in Figure 4. According to (J. M. Linacre, 1994), INFIT and OUTFIT mean square ranging from 0.6 to 1.4 are considered productive for rating scale measurement. Here, the reported mean square values (Infit MNSQ) for Category 2 were the lowest = 0.90; while (Outfit MNSQ) for

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

Category 4 was the highest = 1.07. This demonstrates that the academician's use of this category is productive for measurement.

The structure calibration: *s* was assessed by the rating classification used is true where s-value being the separation among each structure category label.

s 2-3 = -0.63-(-2.39) = 1.76 This value is Ok, meaning that it is acceptable.
s 3-4 = 3.02-(-0.63) = 3.83 This value is Ok, meaning that it is acceptable.

Each response category was calibrated within -2.39, -0.63 and +3.02 *logit*. The trend of the responses can be viewed from the observed average (OBSVD AVRGE) in Figure 4. The values that are normally recommended are a minimum of 1.4 *logit*s (J. M. Linacre, 1994). Figure 2 shows that the order of increment 'Observed Average' was ascending from negative (-0.96 l*ogit*) to positive (2.86 *logit*) value. The calibration of all scales could be seen starting from scales 1, 2, 3 and 4.

SUMMARY OF CATEGORY STRUCTURE. Model="R"

CATEGORY LABEL	SCORE	OBSERVED COUNT	%	OBSVD AVRGE	SAMPLE EXPECT	INFIT MNSQ	OUTFIT MNSQ	ANDRICH THRESHOLD	CATEGORY MEASURE	
1	1	99	3	-.96	-1.00	1.02	1.03	NONE	(-3.60)	1 Strongly Disagree
2	2	615	21	-.23	-.13	.90	.89	-2.39	-1.52	2 Disagree
3	3	1680	57	1.01	.94	.97	1.03	-.63	1.22	3 Agree
4	4	576	19	2.86	2.96	1.08	1.07	3.02	(4.14)	4 Strongly Agree

OBSERVED AVERAGE is mean of measures in category. It is not a parameter estimate.

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	RMSR	ESTIM DISCR	
1	NONE		(-3.60)	-INF	-2.71		83%	5%	1.2934		1 Strongly Disagree
2	-2.39	.11	-1.52	-2.71	-.39	-2.53	54%	43%	.6223	.99	2 Disagree
3	-.63	.05	1.22	-.39	3.09	-.52	69%	84%	.3509	1.10	3 Agree
4	3.02	.06	(4.14)	3.09	+INF	3.05	70%	49%	.7101	.93	4 Strongly Agree

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

Figure 4: Summary of Category Structure

Figure 5 demonstrates how the rating scale was applied in the instrument in concurrence with the model expectations. (Linacre & Wright, 1994) stated that the recommended value for the step calibration structure is a minimum of 1.4 logits. The rating scale was collapsed using the scale (1234). This illustrates that the rating scale is functioning as intended. The value of structure calibration shows the difference between each threshold.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

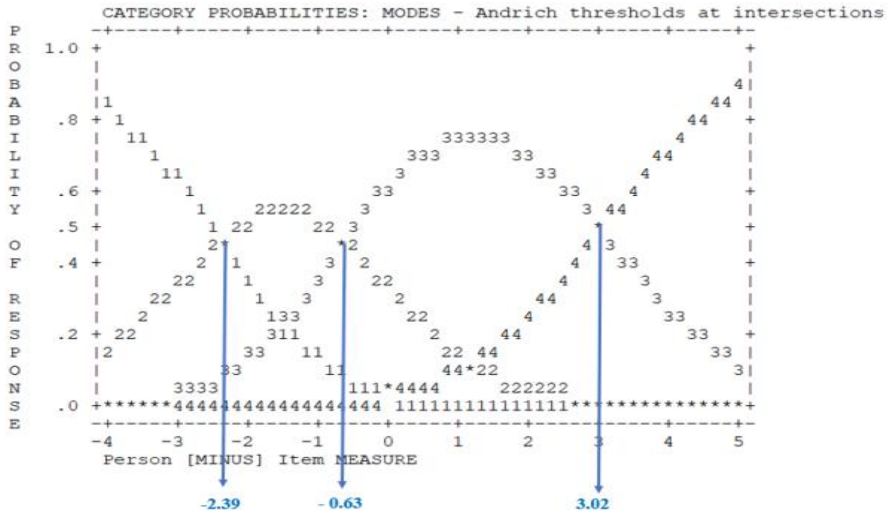

Figure 5: Rating Scale After Cleaning The Data

Another analysis was conducted using Category Probability Curves (CPC) to look at the endorsement of the items and responses. Figure 6 demonstrates the category probability curve of Item 1 (*Ua*) where all the categories function as intended. At this time, in each of the response categories on the Likert scale, there was a region of person locations where that category is the most likely category to be chosen. In these curves, the categories were shown in red, blue, pink and dark grey colour for a Likert scale of 4 levels. The category thresholds occurred in numerical order along the x-axis in the case where an item in which categories function as expected.

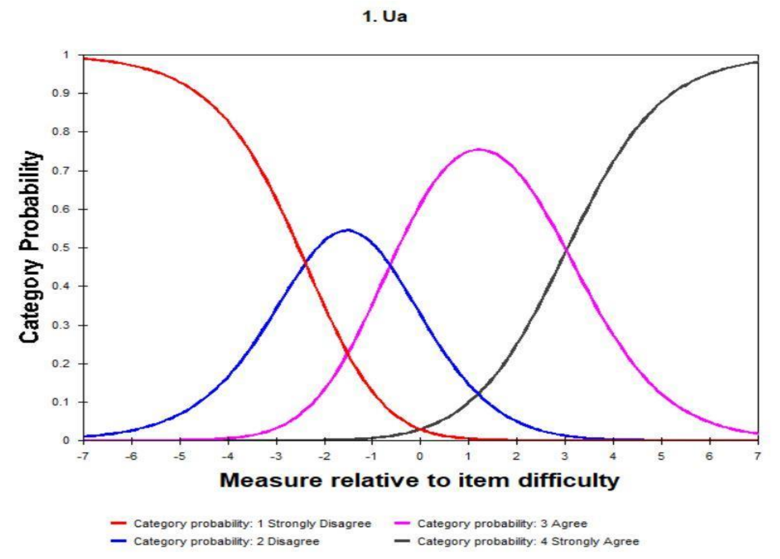

Figure 6: Category Probability Curves (CPC)

To investigate in detail, the largest standardised residual correlation must be identified. Table 3 illustrates the largest Standardised Residual Correlation in identifying dependent items. Based on the value correlation, there was no item with a correlation value of more than 0.70. Items that have a correlation value of more than 0.70 was considered confusing or redundant.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

Table 3: Largest Standardised Residual Correlation Used to Identify Dependent Item

CORRELATION	ENTRY NUMBER ITEM	ENTRY NUMBER ITEM
0.70	13 Ra	14 Rb
0.62	15 Rc	18 Rf
0.59	52 Ia	53 Ib
0.58	37 Sj	39 Sl
0.55	38 Sk	39 Sl
0.54	48 Sec	59 Ih
0.53	53 Ib	54 Ic
0.53	9 Ui	10 Uj
0.52	2 Ub	4 Ud
0.51	53 Ib	57 If
0.51	56 Ie	60 Ii
0.50	14 Rb	17 Re
0.49	55 Id	57 If
0.48	25 Fc	26 Fd
-0.52	30 Sc	57 If
-0.50	17 Re	38 Sk
-0.50	20 Ea	46 Sea
-0.50	5 Ue	31 Sd
-0.47	23 Fa	54 Ic
-0.47	37 Sj	47 Seb

Figure 7 depicts that the raw variance explained by measures was 41.9 per cent, which increased by 7.6 per cent. According to (Fisher, 2007), the value of unexplained variance in 1st contrast between the range of 10 - 15% is considered good. Here, the result for unexplained variance in 1st contrast was 7.7 (8.30%). The situation occurred because there were noise items in the questionnaire. The noise item can be referred to as the same meaning or duplicate question in the survey. Before the data were cleaned, the value was 34.3 per cent. Next, the existence of item dependency showed that there was item dependence that caused noise items. The value was 6.8 per cent, which was considered good.

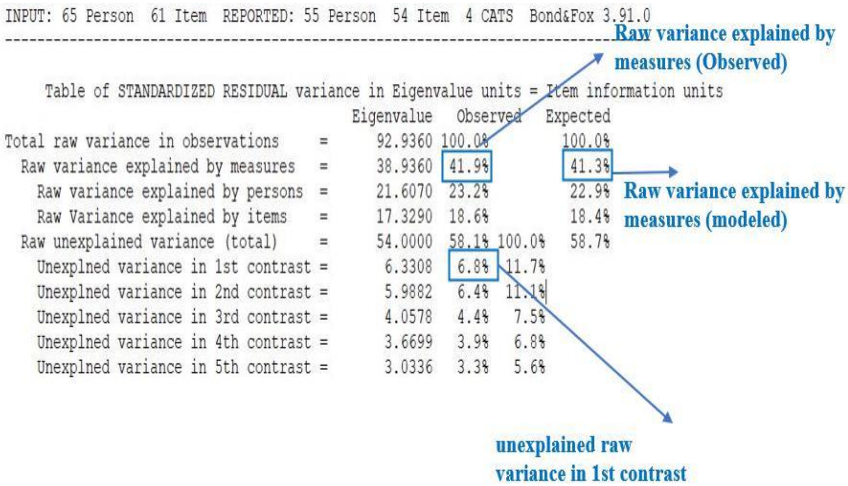

Figure 7: Standard Residual After Cleaning The Data

In conclusion, Unidimensionality was assessed for each of the four subscales. In addition, the quality of the 4-point Likert-type rating scale was assessed by analysing the step calibrations, category fit statistics and category probability curves. In considering all quality indicators simultaneously, it appeared that all four categories were adequately used.

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

Conclusion

This study aims to construct an item questionnaire by using Rasch measurement model by testing the validity and reliability. All 49 item questionnaires have been accepted and validated by Rasch analysis as they meet the criteria (Wahyuningsih, 2020). Rasch is able to analyse and evaluate the item questionnaire by checking the person and item misfit, category capability curve (CPC) and also item calibration. The set of questionnaires produced in this research could be used to conduct other similar research in this domain; however, new questionnaires may be added to suit the research conducted.

Acknowledgement

The authors would like to thank and acknowledge University Selangor for giving an opportunity to conduct this research and to the invited reviewer panel for constructive comment and suggestion.

References

Ariffin, S. R., Omar, B., Isa, A., & Sharif, S. (2010). Validity and reliability Multiple Intelligent item using Rasch measurement model. *Procedia - Social and Behavioral Sciences*, *9*, 729–733. https://doi.org/10.1016/j.sbspro.2010.12.225

Balapumi, R., Konsky, B. R. Von, Mcmeekin, D. A., & Aitken, A. (2015). *Factors Influencing University Students ' Self- Regulation of Learning : An Exploratory Study*. 1–9.

Banawi, A., Rumasoreng, M. I., Basta, I., & Latuconsina, A. (2023). Are the Final Semester Examination Questions for Basic Science Concepts Made by Lecturers Have the Quality According to the Rasch Model? *Jurnal Prima Edukasia*, *11*(1), 81–95. https://doi.org/10.21831/jpe.v11i1.53633

Bilqis, A., Nurhanivah, D., Ramdhani, S. Z., Taufani, M. R., Aminudin, H., Kuncoro, A. K., Syamsiah, E. N., & Ulya, N. (2024). *Identification of Students ' Comprehension of Radiation Concept Using Rasch Identification of Students ' Comprehension of Radiation Concept Using Rasch Analysis*. *July*. https://doi.org/10.46336/ijrcs.v5i3.713

Bond, T. G., & Fox, M. C. (2015). *Applying The Rasch Model - Fundamental Measurement in the Human Sciences*. Routledge. https://doi.org/10.1111/j.1745-3984.2003.tb01103.x

Fisher, W. P. J. (2007). Rating Scale Instrument quality criteria. *Rasch Measurement Transaction*, *21*(1), 1095. http://www.rasch.org/rmt/rmt211a.htm

Linacre, J. . (2009). *A user's guide to WINSTEPS. Chicago, IL: Winsteps*.

Linacre, J. M., & Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Measurements Transactions*, *8*(3), 370.

Mark Wilson. (2005). *Constructing Measures An Item Response Modelling Approach*. Lawrence Erlbaum Associate, Inc.

Mohd Suradi, N. R. (2022). A Process In Developing A Quality Model For Educational Web Application. *Proceedings of the International Conference on Sustainable Practices, Development and Urbanisation (IConsPADU 2021), 16 November 2021, Universiti Selangor (UNISEL), Malaysia*, *3*, 646–653. https://doi.org/10.15405/epms.2022.10.61

Mohd Suradi, N. R., Kahar, S., & Jamaluddin, N. A. A. (2018). Identification of software quality characteristics on academic application in Higher Education Institution (HEI). *Journal of Telecommunication, Electronic and Computer Engineering*, *10*(2–7), 133–136.

Natsir, I., Mayasari, D., & Munfarikhatin, A. (2022). Rasch Model: Quality of Test Instruments to Measure Students' Ability in Algebra. *SHS Web of Conferences*, *149*, 01033. https://doi.org/10.1051/shsconf/202214901033

Journal of Information System and Technology Management
JISTM
EISSN: 0128-1666
Volume 10 Issue 38 (March 2025) PP. 01-11
DOI: 10.35631/JISTM.1038001

Nazurah, M., Shaffiei, Z. A., Daud, N. A., Ahmad, N. D., & Shaffiei, Z. A. (2023). HealthyHeart Data Visualization: Predicting Heart Condition Using Machine Learning. *Journal of Advanced Research in Applied Mechanics*, *105*(1), 41–57. https://doi.org/10.37934/aram.105.1.4157

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*.

Rasch, G. (1960). (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.* Danmarks Paedagogiske Institut, 1960. https://books.google.com.my/books/about/Probabilistic_Models_for_Some_Intelligen .html?id=aB9qLgEACAAJ&redir_esc=y

Singh, P., Fook, C. Y., &, & Sidhu, G. K. (2015). *A Comprehensive Guide in Writing A Research Proposal.* Venton Publishing Sdn Bhd (M).

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch : pada Assessment pendidikan* (2015th ed., Issue October). Trim Komunikata.

Taherdoost, H. (2017). Determining sample size; How to calculate survey sample size. *International Journal of Economics and Management Systems*, *2*(2), 237–239. http://www.iaras.org/iaras/journals/ijems

Taherdoost, H. (2019). What Is the Best Response Scale for Survey and Questionnaire Design; Review of Different Lengths of Rating Scale / Attitude Scale / Likert Scale by Hamed Taherdoost :: SSRN. *International Journal of Academic Research in Management (IJARM)*, *8*(1), 1–10.

Zikmund, W. G., Carr, J. C., Babin, B., & Griffin, M. (2013). *Business Research Methods*.