OPEN ACCESS

Research Paper

DINAMIKA Jurnal Ilmiah Pendidikan Dasar, 2022, 14(2)
ISSN:2655-870X (online)
https://doi.org/10.30595/Dinamika/v14i2.14786

Establishing the Validity and Reliability of a Program Evaluation Questionnaire using Rasch Measurement Model

Su Ling Loh[1], Nor Aisah Jamil[2] & Khar Thoe Ng[3,*]

[1, 3*] Training and Research Division, SEAMEO RECSAM, Penang, MALAYSIA
[2] Port Dickson Education Department, Seremban, MALAYSIA

Received 18 August 2022 ▪ Revised 15 September 2022 ▪ Accepted 30 September 2022

ABSTRACT

This study was conducted to establish the validity and reliability of the evaluation questionnaire for the 12th Regional Congress of the Search For SEAMEO Young Scientist (SSYS) 2022 using the Rasch Measurement Model that was aided by the Winsteps software. The questionnaire contains 24 items that evaluate the Congress's objectives, inputs, as well as event management and administration. Each item is rated on a 4-point rating scale. The instrument was administered at the end of the 3-days SSYS Congress held virtually in which 1891 participants submitted their responses. The establishment of validity and reliability of this questionnaire is crucial before further analysis is carried out. The Rasch Model analysis showed that the reliability index of the respondents was 0.87 and person separation is 2.60, while the item reliability index is 0.96 with an item separation index of 5.08. Item polarity indicates that the point measure correlation (PTMEA CORR) for the 24 items is between 0.67 to 0.76. In terms of item fit, the results indicated one misfit item that needs improvement in the future. The Principal Component Analysis (PCA) shows that almost all the items are unidimensional and intended to measure a similar trait. All these indicate the reliability of the questionnaire, and researchers can proceed with further data analysis to evaluate the 12th Regional Congress of the SSYS 2022.

Keywords: Technological knowledge, Technological Content Knowledge, Technological Pedagogical Knowledge, Science, Augmented Reality

INTRODUCTION

Background and Overview

The 12th Regional Congress of Search for SEAMEO Young Scientists (SSYS) 2022 was a three-dayevent countries organised by the South East Asia Minister of Education (SEAMEO) Regional Education Centre for Science and Mathematics (RECSAM) for youths and educators from the SEAMEO member countries to gather and share information about their scientific and mathematical research projects. It seeks to provide opportunities for youths to optimise their vast potential in conceptualising scientific and mathematical ideas through intellectual teamwork into more tangible research projects and innovative products which will benefit the environment, society, and economy. However, before the further analysis of the questionnaire, reliability and validity need to be established. They are the two most crucial and fundamental criteria in the evaluation of any measurement instrument for successful research. These are crucial elements in contemporary research since they are

✉ lesmatinfodoc@gmail.com **(*Correspondence)**

utilised to improve the accuracy of assessment and evaluation [1]. Hence this study is conducted.

Problem Statement and Methodological Issues

SSYS was initiated in 1997 as biennially held congress since then and were conducted at SEAMEO RECSAM as reported by Ng [2] as well as Mangao and Ng [3] till 2018 as event that was conducted on site [4]. However, SSYS 2020 was cancelled due to pandemic. In 2022, SSYS was organised virtually for the first time due to the current travel restriction caused by the Covid-19 Pandemic. At the end of the third day, an online SSYS evaluation questionnaire in Google Form was disseminated to the participants. It is one of the techniques that can be applied to assess the outcomes of this Congress. The analysis of the questionnaire can be used to decide whether to continue, repeat or improve the programme [5].

Data analysis of this study will be based on the Rasch measurement model. The Rasch measurement model is one of the Item Response Theory models (ITR). ITR states that item features are independent of respondents' ability or proficiency, and respondents' ability or proficiency is independent of the test [6]. Rasch model transforms ordinal response (the Likert scale) into log odd values based on the probability of success, which depends on the differences between person ability and item difficulty. The interaction of log odd values between person ability and item difficulty results in categorising persons based on ability while items are categorised based on difficulty respectively. The categorisation is based on two assumptions: 1. A person with a higher ability is likely to endorse all the items. This contrasts with the characteristics of classical test theory procedures, which rely on respondent samples to determine item difficulty and compare respondents' abilities using the same test items in the same scenario. The logistic distribution defines the item characteristic curve in the Rasch Model, and it has the same shape for all items, with the difficulty of the item being the sole difference [7]. As a result, because it is more than just a statistical description, the Rasch Model provides a remarkably accurate and impartial approach to characterise specific items and people.

The Rasch measurement model generates useful statistics to establish the validity and reliability of an instrument. Validity and reliability are crucial aspects of any instrument or questionnaire to ensure it measures what it should measure and that the result is consistent when repeated. This study aims to establish the reliability and validity of the 12th Congress SSYS questionnaire using the Rasch Measurement Model. The validity of the questionnaire can be established from the perspective of Item and Person Reliability, Item Fit and polarity through the Point Measure Correlation[8]. Besides, Rasch Measurement Model also provides empirical evidence about unidimensionality, which indicates the reliability of a test [9].

MATERIAL AND METHODS

This study is based on an evaluation for the 12th Regional Congress of SSYS using a questionnaire consisting of 24 items rated using the four-point Likert type response, namely 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree. The questionnaire was

converted into Google From and disseminated at the end of the 3rd day through a link posted in the online platforms' chat box, namely Zoom and Youtube live. A total of 1891 participants from different countries responded and submitted their responses. The items in the questionnaire were designed based on the objectives and programs for the Congress listed in the SSYS Programme Booklet [10]. The list of the items in the questionnaire is presented in Appendix 1.

Rasch model analysis was carried out through the application of Winsteps version 3.71.0.1. In determining the validity of an instrument, the Item Polarity, Item Fit, Separation Index, Person Reliability and Item Reliability are used [8]. For Item Polarity, the point measure correlation (PTMEA CORR) must be above 0, and for Item fit, the Mean Square (MNSQ) infit and outfit must be within 0.6 to 1.4, respectively [3]. For the Separation value, all items must show equal or more than 2.0 to be considered good isolation index [8]. Lastly for Item Reliability and Person Reliability, both must have a value of more than 0.8 respectively [6]. Table 1 presents a summary of the statistical criteria for validity and reliability using Rasch Model.

Table 1. Summary of Validity and Reliability Criteria using Rasch Model

Statistical Measurement	**Criteria**
Item Polarity	PTMEA CORR > 0
Item Fit	Infit MNSQ of 0.6 – 1.4 Outfit MNSQ 0.6 – 1.4
Separation	Value ≥ 2.0
Reliability	Value > 0.8

FINDINGS AND DISCUSSION

Reliability and Separation Index

The reliabilities of the Rasch analysis measure the proportion of variance that is true variance for items and person, respectively. Reliability ranges from 0 to 1, with 1 being the best reliability. Person reliability depends on the range of the person's ability, the length of the test, the number of categories per item, and the sample-item targeting. It is independent of the sample size [11]. The item reliability in this study is 0.96 logits, which is very good and effective with a high level of consistency as it is greater than 0.8 logits [6]. The separation index (SI) of item is 5.08 logits (strata = 7) obtained also shows a good isolation index, which is in line with the recommendations by Linacre [12] that state any SI value greater than 2 logit is good. It means the items are evenly distributed and have a high level of reliability on the logits scale. If the items in the questionnaire are sufficiently well separated to define several statistically distinct strata, we are ready to examine their ordering to see whether it makes sense[13]. The person reliability is 0.87, which is also above 0.8 and is good and acceptable. In addition, the SI value is recorded as 2.61 (strata , a good isolation index that is greater than 2.0. The person SE value indicates the degree to which the participants can be differentiated into groups according to their ability. The separation index and strata examine the number of statistically distinct measures. Separation is useful if the persons or items show a normal distribution, and strata are useful if the items or persons are heavy

tailed. Because of the floor-ceiling effect of persons, strata (calculated using person separation) are used in this study to examine how many levels of the quality of the programs for the Congress is able to identify. Item separation is used to verify item hierarchy. The number of strata for items can be calculated through $\frac{4SI+1}{3}$, and it goes the same with the number of strata for persons [13]. A larger sample is required if item reliability and separation are below the required values; if person reliability and separation are below the required values, the test requires more items [14]. Table 2 presents the analysis of the reliability and separation index for item and person. Besides, the Cronbach Alpha value (KR-20) is 0.97, a value greater than 0.8 indicating a high reliability level [6].

Table 2. Reliability and Separation Index for Item and Person

Criteria	Reliability	Separation	Number of strata
Person	0.87	2.61	4
Item	0.97	5.08	7
Cronbach Alpha (KR-20) = 0.97			

Item Polarity

The validity of the items can be referred to the analysis of the program's output, which is the polarity of the item and indicated by the Point Correlation Measure (PTMEA CORR). The polarity of an item determines if it was designed to meet its goals and to measure what should be measured. Examination of polarity of the item is intended to test whether the construct has been built to achieve its objectives. If PTMEA CORR value is greater than 0, it indicates that the item measure what it should measure and also the item is moving in parallel [6]. Conversely, if the PTMEA CORR value is less than 0, it indicates the item does not achieve its intended its objective and should be revised or discarded. The item is out of focus and difficult to be answered by the respondents. Also, this indicates that the response of a person or item conflicts with the variables constructed, an inverse direction of measurement and an uncommon decision-making variable [15], [16]. Table 3 presents the analysis of PTMEA CORR shows all items have positive values between 0.67 to 0.76 and no items show a negative value. Therefore, no items need to be discarded as it has met the minimum requirement. This also indicates that all the items are able to measure what be measured and can distinguish the ability of the respondents. Finlayson and Nunally [17], [18] both believed that the PTMEA CORR item value of at least +0.30 logits would be able to measure a construct systematically, whereas a value of +0.67 logits would be able to merely measure in an average manner. However, this study, the value 0.67 to 0.76 logits in order to prove that the constructed items would be able to be measured and to also be able to differentiate the respondents [15], [16], [19].

Table 3. Item Polarity based on Pont Measure Correlation (PTMEA CORR)

Entry number	Measure	PTMEA CORR
3	-0.38	0.67
4	-0.61	0.67
5	0.31	0.68
11	0.38	0.69
10	0.29	0.69
23	-0.33	0.70
21	-0.45	0.70
17	-0.09	0.70
14	0.73	0.70
7	0.24	0.71
24	-0.75	0.71
20	-0.20	0.71
8	-0.43	0.72
6	-0.41	0.72
19	0.30	0.72
12	0.88	0.73
9	0.97	0.73
16	-0.35	0.73
13	0.29	0.73
1	-0.51	0.74
2	-0.25	0.74
18	0.19	0.75
15	0.11	0.75
22	0.08	0.76

Item Fit

The analysis of item fit estimates the suitability of the items measuring a latent variable. This is represented by the value of the mean square (MNSQ) of Infit and Outfit of each item which should be between 0.6 and 1.4, respectively [6]. If the value is greater than 1.4, the item is considered difficult or confusing. On the other hand, if it shows a value of less than 0.6, it means that the item is too easy [12]. Apart from that, the value of z Infit and Outfit ZSTD (z-Standard) must be between -2.0 and 2.0. However, if the value of Infit and Outfit MNSQ is within the accepted range, then the index of ZSTD is negligible [6]. Therefore, if any item does not meet this requirement, it must be revised or considered for elimination. In analysing item fit, Winsteps recommends a general principle of 1. Investigate outfit before infit; 2. mean-square before t standardised; 3. high values before low or negative

values.

This study showed that the infit MNSQ value for the items ranges from 0.73 to 1.47 while the outfit MNSQ value is between 0.59 to 1.70. There is only one item (Item 5) in which the values are outside the range as the infit MNSQ value is 1.47 exceeding the maximum of 1.4. Similarly, the outfit MNSQ of 1.70 also exceeded the limit value of 1.4. Therefore, item 5 is considered to be revised or eliminated from the questionnaire. Apart from that, item 17 shows an outfit MNSQ value of 0.59, slightly below the minimum of 0.6. However, the infit MNSQ value is recorded as 0.77, thus, this item is maintained in this study. Table 4 presents the summary of item fit based on MNSQ of infit and outfit values.

Table 4. Item Infit and Outfit Mean Square value

Item	Infit MNSQ	Outfit MNSQ
5	1.47	1.70
19	1.29	1.37
21	1.26	1.30
3	1.13	1.19
18	1.09	1.16
23	1.10	1.03
22	1.09	1.08
6	1.08	1.06
20	1.05	1.05
14	1.01	0.99
12	0.96	0.93
9	0.96	0.78
11	0.93	0.76
4	0.92	0.90
16	0.92	0.86
2	0.90	0.85
8	0.89	0.77
24	0.84	0.88
13	0.88	0.81
15	0.88	0.73
10	0.83	0.62
7	0.82	0.82
17	0.77	0.59
1	0.73	0.70

Unidimensionality

Dimensionality refers to the number of latent traits that determine item responses which is one of the key assumption in Item Response Theory (ITR) [20]. Unidimensionality means that all the non-random variance in the data can be accounted for by a single dimension of difficulty and ability [21]. In other words, all items in the instrument or test measure the same latent trait. One of the methods to assess unidimensionality is through principal component analysis (PCA) on residual score, which is the difference between an observed

score and its expected score [20]. The two indicators in the PCA used to assess unidimensionality are the Raw Variance Explained by Measures (RVEM) and Unexplained Variance in First Contrast (UVFC) [6].

In this study, the PCA indicates RVEM is 44.6% (empirical) exceeded the minimum point of 40.0%, as required in the Rasch measurement model, indicating a strong principal measurement dimension [15], [22], [23]. As for UVFC, Linacre (2012) recommended a cut off of 5% and eigenvalue of less than 2.0 to dismiss multidimensionality. However, a value of 5% to 10% and eigenvalue of less than 3.0 are also considered acceptable according to Aziz et al. [21] and Talib et al. [22] respectively. The UVFC for this study is recorded as 5.9% with eigenvalue of 2.6. This indicates that the first contrast also has the strength of the unmodeled variance in about 3 items (eigenvalue 2.6). This is more than the Rasch-predicted chance-value which is usually between 1.5 and 2.0. This seems to indicate a less unidimensional of the instrument and may suggest a multidimension. However, multidimensionality is difficult to define since it is dependent on the instrument's intended function and "Unidimensionality" is a choice based on the circumstances [24]. We cannot split this multidimensional instrument into unidimensional instruments. However, we may decide that this instrument is measuring a general "program evaluation" variable, and declare that this instrument is unidimensional for our purposes. Table 5 presents the findings of RVEM and UVFC.

Table 5. Principal Component Analysis (PCA)

Measures	Value
Raw Variance Explained by Measure (RVEM)	44.6 %
Unexplained variance in First Contrast (UVFC)	5.9 %
Eigenvalue of UVFC	2.6

CONCLUSION

The following conclusions are derived from the research:

1. The developed questionnaire to evaluate the SSY programme is highly reliable, marked by the value of the Cronbach alpha reliability coeffient of 0.97, which is much higher than that of the minimum requirement of 0.7. In addition, the reliability of test items and respondents, and also their separation is also high.
2. The PTMEA CORR values, that measure the item polarity for all the 24 items are within the range of 0.67 to 0.76. Thus, all the items are value and can measure that they should measure.
3. There is only one item is considered misfit, having the values of Infit and Outfit MNSQ beyond the specified limit (0.6 – 1.4). Item 5 which has the Infit MNSQ of 1.47 and Outfit MNSQ of 1.70, need to be revised or considered to be removed.
4. The questionnaire to evaluate SSY programme has good unidimensionality (RVEM = 44.6%, more than 40% and UVFC = 5.9% less than 10% with Eigenvalue of UVFC =

2.6), meaning the instrument measures a single underlying construct (the latent variable) and each item 'fits' the underlying construct.

By using the Rasch model, the results suggest that the evaluation questionnaire for the SSYS programme is valid and reliable to evaluate the effectiveness of the programme with good psychometric performance. Therefore, further analysis can be performed on the questionnaire to obtain more and detailed information about the SSYS programme

REFERENCES

[1] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *Int. J. Med. Educ.*, vol. 2, pp. 53–55, Jun. 2011.

[2] K. T. Ng, "An evaluation of the scientific creativity and problem-solving behaviours of young learners in the development of investigative project work.," in *International Conference on Science and Mathematics Education (CoSMEd)*, 2005, pp. 118–129.

[3] D. D. Mangao and K. T. Ng, "Search for SEAMEO Young Scientists (SSYS) - RECSAM's initiative for promoting public science education: The way forward.," in *International Conference on Science Education 2012*, 2014, pp. 45–56.

[4] K. T. Ng, B. Baharum, M. Othman, S. Tahir, and Y. J. Pang, "Managing technology-enhanced innovation programs: Framework, exemplars and future directions.," *Solid State Technol.*, vol. 63, no. 1, pp. 555–565, 2020.

[5] D. L. Stufflebeam, "The CIPP Model for Evaluation," in *Evaluation Models*, T. Stufflebeam, D. L., Madam, C.F. & Kellaghan, Ed. Boston: Kluwer Academit, 2000, pp. 279–317.

[6] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New Jersey, USA: Lawrence Erlbaum Association, Inc., 2007.

[7] T. Rachman and D. B. Napitupulu, "Rasch Model for Validation a User Acceptance Instrument for Evaluating E-learning System," *CommIT (Communication Inf. Technol. J.*, vol. 11, no. 1, p. 9, 2017.

[8] B. D. Wright and M. H. Stone, *Best Test Design*. Chicago, IL: Mesa Press, 1979.

[9] A. H. Ishak, M. R. Osman, M. H. Mahaiyadin, M. A. Tumiran, and N. Anas, "Examining unidimensionality of psychometric properties via rasch model," *Int. J. Civ. Eng. Technol.*, vol. 9, no. 9, pp. 1462–1467, 2018.

[10] RECSAM, "Programme Book of the 12th Regional Congress of Search for SEAMEO Young Scientists." SEAMEO RECSAM, pp. 1–23, 2022.

[11] S. Moeini, J. V Rasmussen, T. W. Klausen, and S. Brorson, "Rasch analysis of the Western Ontario Osteoarthritis of the Shoulder index - the Danish version," *Patient Relat. Outcome Meas.*, vol. 7, pp. 173–181, Nov. 2016.

[12] J. M. Linacre, *A user's guide to WINSTEPS Rasch-model computer programs*. Chicago, IL: Mesa Press, 2007.

[13] B. D. Wright and G. N. Masters, *Rating Scale Analysis: Rasch Measurement*, vol. 78, no. 382. 1982.

[14] J. M. Linacre and B. D. Wright, "A user's guide to WINSTEPS ministeps Rasch model computer programs." Mesa Press, Chicago, USA, 2012.

[15] J. M. Linacre, *Winsteps Rasch measurement computer program User's guide*. Oregon, USA: Beaverton, 2012.

[16] A. . Azrilah, *Rasch measurement fundamentals: Scale Construct and Measurement Structure.* Kuala Lumpur, Malaysia: Integrated Publishing., 2010.

[17] J. C. Nunnally and I. H. Bernstein, *Psychometric theory*. New York, NY, USA: McGraw-Hill, 1994.

[18] M. L. Finlayson, E. W. Peterson, K. A. Fujimoto, and M. A. Plow, "Raschvalidation of the falls prevention strategies survey," *J. Arch. Phys. Med. Rehabil.*, vol. 90, no. 2, pp. 2039–2046., 2009.

[19] W. P. Fisher, "Rating scale instrument quality criteria," *Rasch Meas. Trans.*, vol. 21, no. 1, p. 1095, 2007.

[20] Y. T. Chou and W. C. Wang, "Checking dimensionality in item response models with principal component analysis on standardized residuals," *Educ. Psychol. Meas.*, vol. 70, no. 5, pp. 717–731, 2010.

[21] J. Sick, "Unidimensionality Equal item discrimination and error due to guessing," *JALT Testing & Evaluation SIG Newsletter.*, vol. 14, no. 2, pp. 23–29, 2010.

[22] A. A. Aziz, M. S. Masodi, and A. Zaharim, *Asas Model Pengukuran Rasch: Pembentukan Skala dan Struktur Pengukuran*. Universiti Kebangsaan Malaysia, 2013.

[23] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Psychology Press, 2013.

[24] J. M. Linacre, "Dimensionality and Structural Validity investigation - an example," 2007. [Online]. Available: https://www.winsteps.com/winman/index.htm?multidimensionality.htm.

APPENDIX

Items in the SSYS evaluation questionnaire

No	Item
	A. Congress Objectives
1	The SSYS Congress has increased student interest in learning science and conducting research project.
2	The SSYS Congress has increased student interest in learning mathematics and conducting research project.
3	The SSYS Congress has provided an opportunity for students to interact socially and intellectually among their fellow student researchers from Southeast Asia
4	The Congress has given recognition to outstanding young scientists and mathematicians in Southeast Asia.
	B. Congress Input
5	The theme "Empowering Innovative Young Minds for Sustainable Future" is very challenging for me and my students to find creative solutions to pressing community, national, regional and global problems for sustainable future.
6	The students' projects presented during the SSYS Congress were of high quality
7	Projects in the SSYS Congress encouraged science learning and research in schools.
8	Projects in the SSYS Congress encouraged mathematics learning and research in schools.
9	The SSYS Congress has increased our awareness on United Nation's 17 Sustainable Development Goals.
10	The SSYS Congress inspired me to be more creative to find solutions to environmental, technological, societal problems including moral and ethical issues.
11	The SSYS Congress inspired me to apply science knowledge and research skills into my daily life, family and community.
12	The SSYS Congress inspired me to apply mathematical knowledge and research skills into my daily life, family and community.
13	The session on 'IoT4Community Remote Desktop for Powering up Projects' was interesting, educational and expanded our knowledge.
14	The games session during the SSYS Congress was fun and interesting.
15	The 'SSYS alumni session' was beneficial to get to know the culture of SSYS delegates with sharing of resources and possible collaboration
16	The atmosphere during the SSYS Congress was lively and conducive for learning.
17	The session on 'Innovation for Sustainable Development' was interesting, educational and increased our knowledge.
	C. Event management and administration
18	Provision of on-line information in SSYS website.
19	Duration of the whole event
20	Presentation of research projects
21	Workshop
22	Icebreaking and SSYS Alumni Session
23	Awards presentation ceremony
24	Overall management of the events