Student Awareness of Space Science: Rasch Model Analysis for Validity and Reliability

Roslinda Rosli[1,*], Mardina Abdullah[2,3], Nur Choiro Siregar[1], Nurul Shazana Abdul Hamid [2,4], Sabirin Abdullah[2], Gan Kok Beng[3], Lilia Halim[5], Noridawaty Mat Daud[2], Siti Aminah Bahari[2], Rosadah Abd Majid[6] & Badariah Bais[3]

[1]Research Centre of Teaching and Learning Innovation, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia

[2]Space Science Centre (ANGKASA), Institute of Climate Change, Universiti Kebangsaan Malaysia, Bangi, Malaysia

[3]Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia

[4]Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

[5]Research Centre of STEM Enculturation, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia

[6]Research Centre of Education and Community Wellbeing, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia

*Correspondence: Research Centre of Teaching and Learning Innovation, Faculty of Education, 43600 UKM Bangi, Malaysia. Tel: 60-3-8921-6290. E-mail: roslinda@ukm.edu.my

Abstract

Validity and reliability are crucial when conducting research to ensure the truthfulness of an instrument. This study investigated the measurement functioning of an instrument on students' awareness of space science. The instrument was administered to 206 secondary school students involved in the Sudden Ionospheric Disturbance π outreach program. Two experts evaluated the content validity of the instrument. Data were analyzed using the Winsteps 3.71.0.1 software to obtain the Rasch model analysis (RMA) on item reliability and persons' separation, item measure, item fit based on PTMEA CORR, polarity items, misfit items, unidimensionality, and a person-item map. The findings revealed that the items are valid, reliable, and appropriate to measure awareness of space science.

Keywords: STEM, questionnaire, Rasch model analysis, Malaysia

1. Introduction

The rapid global development of science and technology has affected the preparation of skillful workers in Science, Technology, Engineering, and Mathematics (STEM) fields (Camilli & Hira, 2019; Waite & McDonald, 2019). In Malaysia, one of the initiatives to boost student interest in STEM fields is the implementation of outreach programs through STEM university grants. The Universiti Kebangsaan Malaysia–Sudden Ionospheric Disturbance π (UKM-SIDπ) program is a platform for secondary school students to engage with space science through hands-on activities (Abdullah et al., 2015; Abdullah et al., 2018; Rosli et al., 2019; Soon et al., 2015; Wah et al., 2012). Activities such as designing the antenna or collecting, analyzing, and verifying data with experts from the university were planned for students to participate actively during the program (Abdullah et al., 2015; Majid et al., 2018). Active involvement through project-based learning could enhance student awareness, creativity, and innovative skills in space science (Abdullah et al., 2015; Abdullah et al., 2018; Majid et al., 2018; Rosli et al., 2019). In conjunction with this program, this study focused on establishing the validity and reliability of the instrument used in the UKM-SIDπ outreach program.

The validity and reliability of an instrument are crucial for any data collection procedure. Validity is the extent to which an instrument measures what it claims to measure and executes what it is designed to measure (Creswell, 2014). There are two types of validity: content validity and construct validity. Content validity refers to the extent to

which an instrument measures the skills/abilities that should be measured (Aiken, 2003; Morrison et al., 2013). Construct validity refers to the readiness of an instrument to assess the concept or content being tested (Sugiyono, 2017). Generally, a group of experts determines the content and construct validity before the pilot test is carried out for factorial analysis. Factorial analysis correlates the item scores and also correlates the factorial scores with the total score (Sugiyono, 2017). Instrument reliability refers to its stability and consistency (Creswell, 2014; Morrison et al., 2013; Sekaran, 2000). In other words, the score obtained from the instrument should be approximately equivalent and consistent when the instrument is frequently tested at different times. This study assessed the validity and reliability of the studied instrument in measuring students' awareness of space science.

2. Method

The present study used a quantitative paradigm to measure the validity and reliability of a questionnaire on student awareness of space science. The questionnaire comprised of 13 items was developed by a group of researchers, who are experts in the field with the help of a few educators. The items were categorized into four sub-scales to (a) space science career, (b) the application of space science, (c) the importance of space science, and (d) the effects of space science in life. A five-point Likert scale was used for students to rate responses (ranging from 1 = *Strongly Disagree* to 4 = *Strongly Agree*, with 5 = *No Opinion*).

Two experts examined the content validity for space science and science education, while construct validity and reliability were established using Rasch model analysis (RMA). Minor changes in wording were made to ensure students understood the content when answering the items. The revised and validated questionnaire was then distributed to several students in a pilot test to ensure the sentence structure of the questions was understandable. The validated questionnaire was then administered to secondary school students (N=206) from four selected schools, who were involved in the hands-on activities in the UKM-SIDπ outreach program. The outreach program began with an educational talk by UKM lecturers in space science, which were followed with small group activities with the assistance of a UKM facilitator (Rosli et al., 2019). The empirical data from the questionnaire were collected before the students participated in the outreach program activities. The data were cleaned before being analyzed using the Winsteps 3.71.0.1 software based on RMA (Linacre, 2011).

3. Results

The validity and reliability of the instrument were analyzed according to the item and person reliability and separation index, item measures, item polarity, misfit items, unidimensionality, and person-item map.

3.1 Item and Person Separation and Reliability

Item and person separation and reliability are a useful measurement in RMA. Table 1 displays the resulting statistical index according to item and person. The separation index indicates the difficulty level of items and persons. Specifically, the person and item separation index indicates that the instrument can separate persons and items into different groups (Bond & Fox, 2015). A suggested value for person and item separation indices is more than 2 (Fox & Jones, 1998). The data analysis here showed person and item separation were 0.89 and 3.50, respectively. The low value for the person separation index (0.89) indicates that the persons are less diverse, as the questionnaire has limited ability to distinguish them. Person and item reliability were 0.44 and 0.92, respectively. According to Fox and Jones (1998), the reliability value is between 0 and 1, while a value greater than 0.8 is acceptable. Here, the item reliability has an extraordinary value of 0.92, which indicates the items are sufficient in terms of size and difficulty to assess the student ability. The person reliability index is weak (0.44), however, which implies that the instrument is not able to differentiate between high and low students. This low value could due to variation in respondent ability, the length of the instrument, several categories per item, and sample-item targeting (Linacre, 2018).

Table 1. Person and Item Separation and Reliability

Criteria	Person	Item
Separation	0.89	3.50
Reliability	0.44	0.92

3.2 Item Difficulty Level

This section discusses the item difficulty level for the questionnaire through logit scale values. There are four categories based on the logit scale: Strongly Agree (less than −1), Agree (−1 to 0), Disagree (0 to 1), and Strongly

Disagree (more than 1; Sumintono & Widhiarso, 2015). Based on Table 2, items I1, I3, I2, I10, and I5 are categorized under difficult to agree with by the respondents (0.1 to 0.94 logit value), while the remaining items (I6, I9, I13, I12, I4, I7, I8, and I11) provoke easy agreement.

3.3 Item Fit Based on Point Measure Correlation

This section shows the extent to which an item fits that which it claims to measure. When an item does not fit, it indicates that respondents were having difficulty understanding the item, and the item should be improved or replaced (Wilmskoetter et al., 2019). The Point Measure Correlation (PTMEA CORR) value can be observed to examine item fit (Boone et al., 2014). The PTMEA CORR index for item category is extremely good for values more than 0.40, good for values between 0.30 and 0.39, and fair for values between 0.20 and 0.29. If the values fall between 0 and 0.19, the questionnaire fails to discriminate items. For a survey in which the PTMEA CORR value is less than or equal to 0, the researchers have to reevaluate the items (Alagumalai et al., 2005). Table 3 presents the PTMEA CORR index for this study. The appropriateness of the questionnaire based on the measured construct falls under the "extremely good" category for items I12, I5, I9, I10, I4, I6, and I8. Items I11, I3, I13, I2, and I7, are classified under the good category, meanwhile item I1 is in the fair category of appropriateness. Based on these PTMEA CORR indices, all of the items can be used to measure student awareness of space science.

Table 2. Item Difficulty Level

Logit scale value	Item	Category
0.94	I1	Disagree
0.87	I3	Disagree
0.20	I2	Disagree
0.12	I10	Disagree
0.10	I5	Disagree
-0.08	I6	Agree
-0.11	I9	Agree
-0.15	I13	Agree
-0.19	I12	Agree
-0.31	I4	Agree
-0.41	I7	Agree
-0.41	I8	Agree
-0.58	I11	Agree

Table 3. PTMEA CORR Index

PTMEA CORR	Item	Category
0.39	I11	Good
0.44	I12	Extremely good
0.54	I5	Extremely good
0.34	I3	Good
0.42	I9	Extremely good
0.29	I1	Fair
0.45	I10	Extremely good
0.31	I13	Good
0.41	I4	Extremely good
0.47	I6	Extremely good
0.38	I2	Good
0.36	I7	Good
0.44	I8	Extremely good

3.4 Item Polarity

The identification of item polarity examines the appropriateness of the developed construct in achieving its objective based on the PTMEA CORR value. If the PTMEA CORR is positive (+), the item measures the construct, and if it is negative (−) or zero, then the item does not measure the construct (Bond & Fox, 2015; Linacre, 2018), which would mean that the researcher needs to either revise or eliminate the items. Referring to Table 4, the item polarity of the

instrument show positive PTMEA CORR values ranging from 0.29 and 0.54, which indicates that each item is valid to measure student awareness of space science.

3.5 Misfit Items

The appropriateness of items in measuring the construct can be observed through outfit and infit Mean-Square (MNSQ) values. MNSQ values should be between 0.6 and 1.4 to ensure item suitability in measuring the construct (Bond & Fox, 2015). If the MNSQ index value exceeds 1.4; therefore, the item is a misfit, while an index value less than 0.6 indicates students easily predict the item. Z scores or Z Standardized (ZSTD) are also essential for finding misfit items, and the value should be from −2 to +2 (Bond & Fox, 2015). When the MNSQ value is accepted, the ZSTD can be ignored (Linacre, 2018). If the MNSQ condition is not fulfilled, the items should either be refined or removed. Table 5 shows the misfit items in the study.

Analysis of Table 5 indicates the infit MNSQ value is from 0.40 to 3.12, while the outfit MNSQ is from 0.38 to 3.12. There are five items outside of the infit MNSQ range. One item, I11, is above 1.4 with 3.12, and four items are below 0.6: I6 (0.59), I2 (0.57), I7 (0.45), and I8 (0.40). Based on MNSQ value, these five items are out of range. The researchers decided to refine these five items (I11, I6, I2, I7, and I8), as the infit and outfit MNSQ score for each item is outside the range of acceptability. The refinement of the items was done to the language and scientific terms used.

Table 4. Item Polarity

Item	Measure	INFIT		OUTFIT		PTMEA CORR
		MNSQ	ZSTD	MNSQ	ZSTD	
I1	.94	.87	-1.1	.87	-1.2	.29
I13	-.15	.84	-1.6	.80	-2.1	.31
I3	.87	1.15	1.3	1.16	1.4	.34
I7	-.41	1.45	-7.1	.45	-7.1	.36
I2	.20	1.57	-5.0	.56	-5.1	.38
I11	-.58	3.12	9.9	3.12	9.9	.39
I4	-.31	.63	-4.3	.63	-4.3	.41
I9	-.11	1.00	.1	1.01	.1	.42
I8	-.41	.40	-8.0	.38	-8.4	.44
I12	-.19	1.38	3.5	1.33	3.1	.44
I10	.12	.86	-1.4	.86	-1.4	.45
I6	-.08	.59	-4.7	.59	-4.8	.47
I5	.10	1.17	1.6	1.17	1.6	.54

Table 5. Misfit Items According to Outfit and Infit MNSQ

Item	Measure	INFIT		OUTFIT		PTMEA CORR
		MNSQ	ZSTD	MNSQ	ZSTD	
I11	-.58	3.12	9.9	3.12	9.9	.39
I12	-.19	1.38	3.5	1.33	3.1	.44
I5	.10	1.17	1.6	1.17	1.6	.54
I3	.87	1.15	1.3	1.16	1.4	.34
I9	-.11	1.00	.1	1.01	.1	.42
I1	.94	.87	-1.1	.87	-1.2	.29
I10	.12	.86	-1.4	.86	-1.4	.45
I13	-.15	.84	-1.6	.80	-2.1	.31
I4	-.31	.63	-4.3	.63	-4.3	.41
I6	-.08	.59	-4.7	.59	-4.8	.47
I2	.20	.57	-5.0	.56	-5.1	.38
I7	-.41	.45	-7.1	.45	-7.1	.36
I8	-.41	.40	-8.0	.38	-8.4	.44

3.6 Unidimensionality

The unidimensional index measures the instrument based on Residual Principal Component Analysis (PCA). Instruments must fit the condition of unidimensionality. If unidimensionality is not fulfilled, it is unnecessary to count the aggregate total and compare respondents or items (Boone et al., 2014). According to Linacre (2018), the minimum variance to show the unidimensionality of the instrument is 20%. The unexplained variance in the first contrast should not exceed 15% (Fisher, 2007). Table 6 shows that the raw variance explained by the measures was 20.7%, which was 0.2% more than the modeled percentage for minimum unidimensionality. All of the unexplained variance in the first, second, third, fourth, and fifth contrasts did not exceed 15%. The given instrument is thus reliable in measuring the construct of student awareness of space science. A unidimensional construct measures what it claims to measure (no presence of other variance outside of the construct).

Table 6. Standardized Residual Variance

	Empirical			Modeled (%)
	Eigenvalue units	%	Variance unexplained (%)	
Total raw variance in observations	16.4	100.0		100.0
Raw variance explained by measures	3.4	20.7		20.9
Raw variance explained by persons	.9	5.8		5.8
Raw variance explained by items	2.4	14.9		15.0
Raw unexplained variance (total)	13.0	79.3	100.0	79.1
Unexplained variance in 1st contrast	2.1	12.5	15.8	
Unexplained variance in 2st contrast	1.5	9.1	11.5	
Unexplained variance in 3st contrast	1.4	8.7	10.9	
Unexplained variance in 4st contrast	1.2	7.4	9.4	
Unexplained variance in 5st contrast	1.1	6.7	8.5	

Standardized residual correlation was examined to check item dependency. An item is dependent and is not single if the standardized residual correlation value is above 0.7 (Linacre, 2018). Table 7 indicates that the standardized residual correlation for all items is less than 0.7, which shows that the items are independent and single. However, based on the outfit and infit indexes, an MNSQ approaching 1.0 must be retained (refer to Table 5).

Table 7. Standardized Residual Correlation

Correlation	Entry Number	Item	Entry Number	Item
.39	7	I7	8	I8
.27	2	I2	8	I8
-.30	10	I10	11	I11
-.30	8	I8	11	I11
-.28	11	I11	13	I13
-.28	4	I4	11	I11
-.24	2	I2	12	I12
-.24	6	I6	12	I12
-.23	7	I7	11	I11
-.20	2	I2	11	I11

3.7 Person-Item Map

RMA identifies the correlation between student ability and item difficulty using a logarithmic function. Student achievement in problem-solving is highly related to student ability and item difficulty (Englehard, 2013; Stelmack et al., 2004). Figure 1 shows the Person-Item Map (PIM), which indicates student ability and item difficulty in the logit scale ranging from −2 to 4. The top left shows students who lack awareness of space science, while the bottom left indicates those who have high awareness. Item difficulty is shown on the right side of the PIM. Students seldom answer "agree" to I11 ("The intensity of the sunlight disrupts telecommunication signal"), while they easily answer "agree" to I1 ("There are a lot of careers related to space science"). Some items (I10, I12, I15, I6, I12, I13, I9, I4, I7, and I8) measure the same construct.
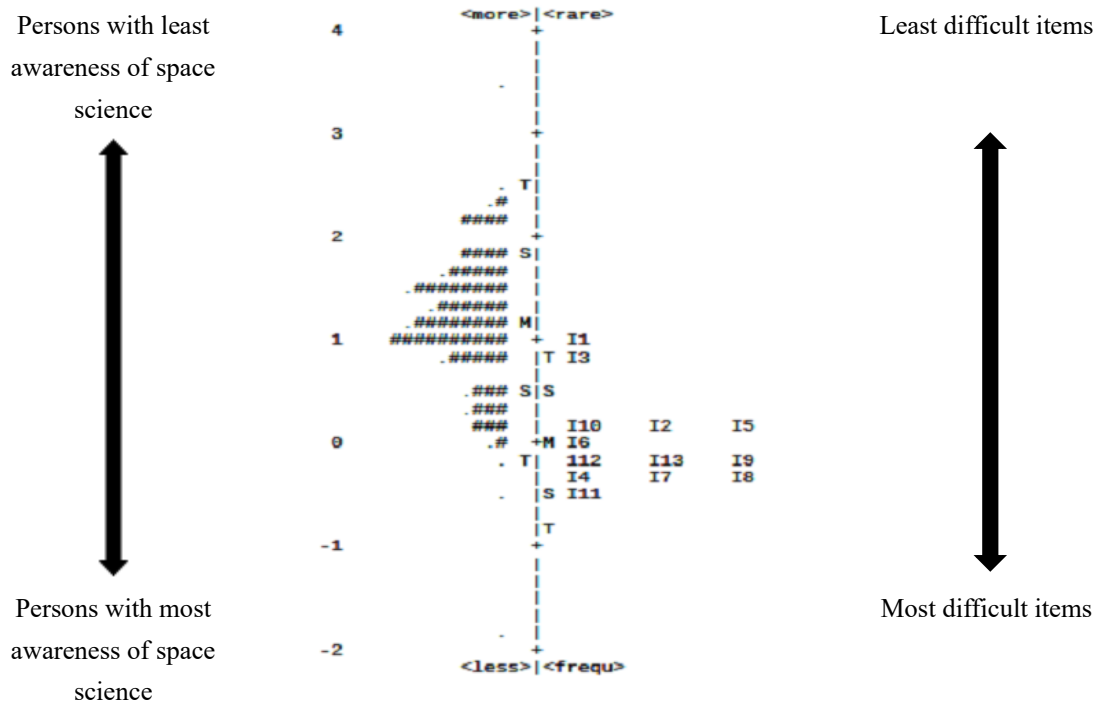
Figure 1. Person-Item Map

4. Discussion

In Malaysia, the level of student awareness of space science is still modest. Through increased STEM education in most countries globally, teachers and university lecturers alike are concerned with developing science education and educating students about the importance of science in the future (Abdullah et al., 2018; Camilli & Hira, 2019; Waite & McDonald, 2019). The UKM-SIDπ program is a successful activity for high school students, which increases student interest in learning science (Abdullah et al., 2015; Abdullah et al., 2018). The questionnaire created in this study is one way to determine how students respond to the subject of space science in a program designed by educators and lecturers to raise awareness among future students.

In the present study, six aspects were analyzed with RMA to show that the questionnaire's assessment of student awareness of space science is valid and reliable. First, the item and person reliability and separation indices are greater than 0.8 (Fox & Jones, 1998). Second, analysis of item difficulty level based on logit value (0.1 to 0.94) revealed five items were difficult to agree with, and eight items were easy to agree with (Sumintono & Widhiarso, 2015). The fit for item values based on PTMEA CORR analyzed three aspects; there are seven items in the outstanding category, five items in the good category, and one item in the fair category (Boone et al., 2014). Fourth, the item polarity of this instrument is between 0.29 and 0.54, which is within the range of validity (Bond & Fox, 2015; Linacre, 2018). Fifth, item misfit was identified, including infit MNSQ values (0.40 to 3.12) and outfit MNSQ values (0.38 to 3.12; Bond & Fox, 2015). Sixth, residual PCA confirmed the unidimensionality of the instrument at 20.9% (Linacre, 2018). Finally, the PIM was placed on a logit scale ranging from −2 to 4. The value of the PIM lies in how it assesses students' ability to cope with item difficulty (Englehard, 2013; Stelmack et al., 2004). Based on the results using RMA, the questionnaire meets the standards of validity and reliability and can be used by teachers or school administrators to determine the level of student awareness of the subject of space science.

5. Conclusion

The content validity of the instrument was carried out by two experts. Based on RMA, construct validity and reliability were also established. The results showed that the questionnaire to measure student awareness of space science is valid and reliable because the items are fitted with what the instrument is supposed to measure. However, the misleading items should be revised or omitted. Future researchers could include more items with non-overlapping descriptions that giving similar meanings or interpretations.

6. Implications

This study proposed a valid and reliable instrument for measuring student awareness of space science. This study thus has implications for teachers and school administrators, who can use the instrument to assess student knowledge in this area. From the results of the analysis conducted, the students can provide feedback allowing teachers to create lessons that will enhance student awareness, especially on the subject of space science.

Acknowledgments

References

Abdullah, M., Bahari S. M., Bais, B., Hasbi, A. M., Majid, R. A., Mokhtar, M. H., & Syaidah, N. (2015). Space weather innovation competition for school students in Malaysia. *International Journal of Humanities, Arts and Social Sciences*, *1*(2), 70-74.

Abdullah, M., Majid, R. A., Bais, B., Bahri, N. S., & Mokhtar, M. H. (2018). Space weather monitoring system competition for secondary school students in Malaysia. In W. Suparta, M. Abdullah, & M. Ismail (Eds.), *Space Science and Communication for Sustainability* (pp. 135-146). Singapore: Springer.

Aiken, S. (2003). Estimation of item parameters. In G. H. Fischer, & I. W. Molenaar (Eds.), *Rasch models, foundations, recent developments, and applications* (pp. 39-52.). New York, NY: Springer-Verlag.

Alagumalai, S., Curtis, D. D., & Hungi, N. (2005). *Applied Rasch measurement: A book of exemplars.* Dordrecht, The Netherlands: Springer.

Bond, T. G., & Fox, C. (2015). *Applying the Rasch model fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht, The Netherlands: Springer.

Camilli, G., & Hira, R. (2019). Introduction to special Issue-STEM workforce: STEM education and the post-scientific society. *Journal of Science Education and Technology*, *28*(1), 1-8. https://doi.org/10.1007/s10956-018-9759-8

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approach* (4th ed.). Thousand Oaks, CA: Sage publications.

Englehard, G. (2013). *Invariant Measurement: Using Rasch models in the social, behavioral, and health sciences.* New York, NY: Routledge.

Fisher, W. P., Jr. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21*(1), 1095.

Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology*, *45*(1), 30. https://doi.org/10.1037/0022-0167.45.1.30

Linacre, J. M. (2011). *Winsteps®* (Version 3.71.0.1) [Computer Software]. Beaverton, OR: Winsteps.com.

Linacre, J. M. (2018). *A user's guide to Winstep mini step Rasch-Model computer programs*. Beaverton, OR: Winsteps.

Majid, R. A., Abdullah, M., Bais, B., Bahri, N. S., Mokhtar, M. H., & Bahari, S. A. (2018). Relationship between students' attitudes in space science with their achievements in the field. *Creative Education*, *9*, 2291-2299. https://doi.org/10.4236/ce.2018.914169

Morrison, G. R., Ross, S. M., Kalman, H. K., & Kemp, J. E. (2013). *Designing effective instruction* (7th ed.). New York, NY: John Wiley & Sons, Inc.

Rosli, R., Abdullah, M., Siregar, N. C., Abdul Hamid, N. S., Abdullah, S., Gan, K. B., Halim, L., Mat Daud, N., Bahari, S. A., Abdul Majid, R., & Bais, B. (2019). Exploring space science through the UKM-SIDπ outreach program. *Proceedings of 2019 6th International Conference on Space Science and Communication* (*IconSpace*) (pp. 253-256). https://doi.org/10.1109/IconSpace.2019.8905957

Sekaran, U. (2000). *Research methods for business: A skill-building approach* (7th ed.). New York, NY: John Wiley & Sons, Inc.

Soon, Y. L., Gan, K. B., & Abdullah, M. (2015, August). Development of very low frequency (VLF) data acquisition system using Raspberry Pi. *Proceedings of the 2015 2nd International Conference on Space Science and Communication (IconSpace)* (pp. 485-488). https://doi.org/10.1109/IconSpace.2015.7283807

Stelmack, J., Szlyk, J. P., Stelmack, T., Babcock-Parziale, J., Demers-Turco, P., Williams, R. T., & Massof, R. W. (2004). Use of Rasch person-item map in exploratory data analysis: A clinical perspective. *Journal of Rehabilitation Research & Development*, *41*(2), 233-241. https://doi.org/10.1682/jrrd.2004.02.0233

Sugiyono. (2017). *Metode penelitian kuantitatif, kualitatif, dan R & D*. Bandung, Indonesia: Alfabeta, CV.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan.* Cimahi, Indonesia: Trim Komunikata.

Wah, W. P., Abdullah, M., Hasbi, A. M., & Bahari, S. A. (2012, December). Development of a VLF receiver system for Sudden Ionospheric Disturbances (SID) detection. *Proceedings of the 2012 IEEE Asia-Pacific Conference on Applied Electromagnetics (APACE)* (pp. 98-103). https://doi.org/10.1109/APACE.2012.6457640.

Waite, A. M., & McDonald, K. S. (2019). Exploring challenges and solutions facing STEM careers in the 21st century: A human resource development perspective. *Advances in Developing Human Resources*, *21*(1), 3-15. https://doi.org/10.1177/1523422318814482

Wilmskoetter, J., Bonilha, H., Hong, I., Hazelwood, R. J., Martin-Harris, B., & Velozo, C. (2019). Construct validity of the eating assessment tool (EAT-10). *Disability and Rehabilitation*, *41*(5), 549-559. https://doi.org/10.1080/09638288.2017.1398787