ISSN 1392-0340 (Print)
ISSN 2029-0551 (Online)
https://doi.org/10.15823/p.2024.153.9

Pedagogika / Pedagogy
2024, t. 153, Nr. 1, p. 185–198 / Vol. 153, No. 1, pp. 185–198, 2024

VYTAUTO DIDŽIOJO UNIVERSITETO ŠVIETIMO AKADEMIJA

Learning Environment of Numeracy: Development and Validation With Confirmatory Factor Analysis and Rasch Model

Cynthia Mayang Sari[1], Achmad Ridwan, Ridwan[2], Yuliatri Sastrawijaya, Sastrawijaya[3]

1 State University of Jakarta, Departament of Education Research and Evaluation, 11 Rawamangun Muka St., Rawamangun, Pulogadung, ID-13220 East Jakarta, Indonesia, cynthia.mayang20@gmail.com

2 State University of Jakarta, Department of Educational Research and Evaluation, Department of Chemistry Education, Indonesia, 11 Rawamangun Muka Raya St., Rawamangun, Pulogadung, ID-13220 East Jakarta, Indonesia, achmadridwan@unj.ac.id

3 State University of Jakarta, Department of Educational Research and Evaluation, Department of Informatics and Computer Engineering Education, 11 Rawamangun Muka Raya St., Rawamangun, Pulogadung, ID-13220 East Jakarta, Indonesia, yuliatri@unj.ac.id

Annotation. This study aims to develop and validate an instrument that measures the learning environment of numeracy. The research was conducted in six junior high schools in Central Kalimantan Province in Indonesia. Analysis was conducted using confirmatory factor analysis (CFA) and Rasch model. The results of the analysis showed that 58 items fit the Rasch model and had item and person reliability of 0.99 and 0.98. The instrument has strong psychometric qualities and a decent factor structure.

Keywords: *numeracy, learning environment, learning environment of numeracy, Rasch model.*

Introduction

The learning environment is one part of the learning process in achieving learning goals, where the learning environment affects teaching and learning activities at school. The setting or context that makes learning possible is known as the learning environment. In this setting, students learn academic material and gain social experience (Bisri et al., 2023). The learning environment is closely related to psychosocial aspects and consists of the psychological, social, cultural and physical environment in which learning occurs (Rusticus et al., 2023). The numeracy learning environment is a learning environment that supports students in developing numeracy skills where students use mathematical thinking in solving everyday problems to be better prepared to face life's challenges (Çelik & Özdemir, 2020).

In a numeracy learning environment, individuals are trained to develop logical thinking and reasoning strategies in everyday activities where numeracy skills are needed to solve problems and understand numbers, time, patterns and shapes for activities such as cooking, reading receipts, reading instructions and even playing sports because everyday life cannot be separated from the relationship with numbers and even math concepts (Insorio & Librada, 2020). Thus, students as individuals not only have numeracy mastery limited to mathematics material but will also be able to formulate, use, and interpret mathematics in various contexts. The numeracy learning environment is one of the important factors that are important and influential in improving students' numeracy skills. This is in line with Hassan and Omar's statement that the learning environment plays an important role in learning activities that affect student learning outcomes (Md Hassan et al., 2020; Omar & Awang, 2023).

As far as researchers have found, there is a gap in research on numeracy learning environments. This can be seen from the lack of literature articles that discuss the issue. Some of the research conducted is more focused on improving numeracy skills through various methods such as those conducted by Cahyani (2023), Bulijan (2019) and Garcia (2019). While research on learning environment is conducted on other contexts or topics such as measurement of learning environment in classrooms by Aluri & Fraser (2019), virtual learning environment measurement by Sole et al. (2022), Aluja et al. (2019) and measurement of the laboratory learning environment by Nikolic et al. (2021).

Assessment of the numeracy learning environment can be realized through appropriate instruments. The resulting instrument must be a valid and reliable instrument in order to provide accurate data in the form of psychometric characteristics that will provide results on how the learning environment is needed to improve student numeracy skills. Instruments that have so far been produced related to the learning environment include *Learning Environment Inventory* (LEI), *Classroom Environment Scale* (CES), *Science Laboratory Environment Inventory* (SLEI), *Constructivist Learning Environement Survey* (CLES), dan *What Happened in The Class* (WIHIC) (Koul, 2023). However, there is a research gap in that no instrument has been found that measures the numeracy learning environment. This research aims to produce a valid and reliable instrument that assesses the numeracy learning environment. In realizing the purpose of this study, the researcher is interested in using combination of confirmatory factor analysis (CFA) and Rasch model is considered necessary to provide comprehensive information on psychometric characteristics that is more in-depth and complementary between the two to produce validity and reliability of the instrument Questionnaire Learning Environment of Numeracy (QLEN).

Research Methodology

Sample

The sample in this study consists of a sample from a limited trial or pilot study and from extensive stage. In the pilot study, Andrade suggested a sample size of between 25 and 100 people (Andrade, 2020). While in extensive stage, the number of samples is recommended to be five to ten times the instrument items (Memon et al., 2020). In this study, the researcher used 264 students in the pilot study and in the extensive stage consisted of 566 students from six junior high schools in Central Kalimantan Province in Indonesia.

Instruments and Procedures

Instrument QLEN consisted of six dimensions, 12 indicators, and 60 items. The instrument provides five response options using a Likert scale with intervals ranging from 1=strongly disagree and 5=strongly agree. This instrument was developed based on the dimensions of the Constructivist Oriented Learning Survey (COLES) instrument. The COLES instrument is an instrument that has been widely discussed and used in many places and (Aluri & Fraser, 2019). The descriptions of the six dimensions as well as sample items from each dimension are described in the following table.

Table 1
Dimensions, Descriptions and Examples of Scale

Dimensions	Description	Examples of Scale
Student Cohesiveness (SC)	The extent to which students know, help and support each other (Aldridge et al., 2012).	I know my friends in class and work well together.
Teacher Support (TS)	The extent to which the teacher helps, interacts with, and gives confidence to students (Dorman et al., 2006).	The teacher relates the subject matter to examples in daily life.
Involvement (IN)	The extent to which students feel that they have the opportunity to participate in discussions and have interest and attention to what is happening in the classroom (Aldridge et al., 2012).	I convey ideas and opinions in class discussions.
Task Orientation (TO)	The extent to which students consider it important to complete activities and understand course objectives (Aldridge et al., 2012).	The teacher gives math problems related to daily life.

Dimensions	Description	Examples of Scale
Autonomy (AU)	The extent to which students have opportunities to initiate ideas and make their own learning decisions, and the locus of control are student oriented (Aldridge et al., 2012; Walker, 2003).	I relate math to everyday life, for example when calculating discount prices, measuring the height and weight of objects, and so on.
Equity (EQ)	The extent to which students receive the same treatment from teachers as other students (Aldridge et al., 2012).	The teacher provides an opportunity for all students to ask questions if there is material that is not understood.

The instrument development procedure in this study used the Recker & Rosemann (2010) development procedure which consists of five development steps. The development step begins with item creation through literature review to determine the constructs and dimensions used and the items arranged in the instrument. The next step is substrata identification, where the selected experts are determined to assess the accuracy of the content validity of the instrument developed. After that, it continued with the item identification step, namely testing the validity of the instrument through expert and panelist validation. The fourth step in the form of item revision aims to determine the measurement items that are most likely to be well understood in the final field test, and the last step is instrument validation in the form of presenting the psychometric characteristics of the instrument in the form of validity and reliability test results (Recker & Rosemann, 2010).

To test the construct validity of the instrument in the pilot study, researchers used confirmatory factor analysis (CFA), a model that assesses the correlation between latent variables and unobserved variables. With the use of the standardized factor loading value, the link between the latent and observable variables is examined. With the aid of the statistical program Lisrel 8.80, the loading factor and goodness-of-fit (GOF) values were calculated using the maximum likelihood model. The GOF is the standard index used to test model fit. The goodness of fit (GOF) metric indicates how well the model reproduces the covariance matrix found in the items (Hair et al., 2010). In recent decades, several metrics have been proposed to evaluate model adequacy, including ratio values χ^2/df, adjusted goodness of fit index (AGFI), comparative fit index (CFI), and root mean square error of approximation (RMSEA) (Ayu et al., 2023).

While in the broad stage of data testing, researchers used the Rasch model in the data analysis carried out. As a modern measurement theory, the Rasch model has several advantages including ordinal data from questionnaires can be converted into interval data using the Rasch model analysis (Mustafa et al., 2021). This approach can reveal people's behavior by simulating quantifiable objects. Indicators of instrument reliability include item reliability, person reliability coefficient, and Cronbach's alpha coefficient. The suitability of the measurement model was evaluated by the outfit MNSQ, outfit ZSTD, and

Pt Mean Corr statistical values. A model fit interval of $0.5 < MNSQ < 1.5$, $-2.0 < ZSTD < +2.0$ and $0.4 <$ Pt Mean Corr < 0.85 is required (Davis & Boone, 2021, p. 2) The degree of internal consistency is ascertained using the Cronbach alpha coefficient. Bias in items is found using differential item functioning, or DIF. The item is biased toward the qualities of the responder if the probability value is less than 5 percent (Ayu et al., 2023).

Results

Confirmatory Factor Analysis

Confirmatory factor analysis resulted that QLEN factor structure's compatibility with the model being examined based on the GOF index with $\chi2/df = 3.14$, CFI = 0,96, AGFI = 0.86, RMSEA = 0.072 and p-value = 0.00. Table 1 displays the findings of the comprehensive model fit analysis. To fulfill the cut-off values, all indices are employed, based on the fit index values in table 2. According to the study's findings, the QLEN scale and the model fit each other well.

Table 2
Goodness of Fit (GOF) of QLEN

	χ2/df	CFI	AGFI	RMSEA	p-value
Cut off	≤5.00	≥0.95	>0.80	<0.08	<0.05
Result	5367.97/1710 = 3.14	0.96	0.86	0.072	0.00

According to Table 3, item loading on QLEN scale ranges from 0.28 to 0.71. The correlation between items and their factors is represented by item or factor loading. Which items on the QLEN scale are valid is determined by this value. If an item's loading value is >0.30, it is regarded as valid item (Pergo et al., 2023), whereas the least level for structural interpretation is satisfied by a loading factor in the range of 0.30–0.40 (Rokhman et al., 2023, p. 249). Item EQ9 "The teacher invites anyone who wants to answer the questions in front of the class" was eliminated because it had a loading factor value smaller than 0.30 and not included in the next stage of analysis. At the significance level of $p < 0.05$, the other 59 items showed good correlations with each factor. Based on the GOF index, the applicability of the QLEN factor structure to the model was examined. Table 3 presents the findings of the loading factor value of QLEN.

Table 3
Loading Factor Value of QLEN

Item	Loading factor	Item	Loading factor	Item	Loading factor	Item	Loading factor	Item	Loading factor	Item	Loading factor
SC1	0.38	TS1	0.35	EQ1	0.42	TO1	0.49	AT1	0.62	IN1	0.71
SC2	0.38	TS2	0.45	EQ2	0.31	TO2	0.48	AT2	0.61	IN2	0.59
SC3	0.32	TS3	0.43	EQ3	0.49	TO3	0.36	AT3	0.60	IN3	0.62
SC4	0.39	TS4	0.45	EQ4	0.34	TO4	0.57	AT4	0.41	IN4	0.70
SC5	0.40	TS5	0.33	EQ5	0.38	TO5	0.69	AT5	0.64	IN5	0.66
SC6	0.33	TS6	0.42	EQ6	0.39	TO6	0.53	AT6	0.58	IN6	0.67
SC7	0.36	TS7	0.43	EQ7	0.55	TO7	0.57	AT7	0.65	IN7	0.53
SC8	0.49	TS8	0.44	EQ8	0.45	TO8	0.43	AT8	0.57	IN8	0.63
SC9	0.47	TS9	0.44	EQ9	0.28	TO9	0.69	AT9	0.62	IN9	0.44
SC10	0.43	TS10	0.42	EQ10	0.40	TO10	0.53	AT10	0.67	IN10	0.47

Rasch Model Analysis

Dimensionality

Testing for unidimensionality is done through principal component analysis (PCA) of the residuals used. In a measure of unidimensionality, it is expected that the observed variance explained by the measure roughly matches the variance expected in the model. The "first contrast" is the component that explains the largest possible amount of variance in the residuals. If the unexplained variance found in the first contrast reaches a 2.0 Eigenvalue, the largest possible secondary dimension has a power of less than 2 items. The raw variance explained by the size of the minimum acceptable value is 40%, while the unexplained variance in the first factor should not be more than 15% (Boone & Staver, 2020, p. 4).

PCA analysis of standardized residual variance shows the raw variance explained by the measurement is 54.4% with unexplained variance at first contrast of 1.3%. These results meet the requirements for a satisfactory measure, indicating that more than 40% of the raw variance is explained by the measurement and less than 15% of the variance is unexplained at first contrast (Ilias & Siew, 2022, p. 346). The fulfillment of this criterion indicates that the assumption of unidimensionality on this instrument has been met.

Table 4

Standardized Residual Variance (in Eigenvalue Units)

			Empirical		Modeled
Total raw variance in observations	=	129.3	100%		100%
Raw variance explained by measures	=	70.3	**54.4%**		52.5%
Raw variance explained by persons	=	31.7	24.5%		23.7%
Raw variance explained by items	=	38.5	29.8%		28.8%
Raw unexplained variance (total)	=	59.0	45.6%	100%	47.5%
Unexplained variance in 1st contrast	=	1.7	**1.3%**	2.9%	
Unexplained variance in 2nd contrast	=	1.7	1.3%	2.9%	

Item Fit

This step is to analyze the data to produce the outfit MNSQ, outfit ZSTD, and Pt Mean Corr values of each instrument item. Fit items are items that meet the criteria of 0.5 < MNSQ < 1.5, -2.0 < ZSTD < +2.0 and 0.4 < Pt Mean Corr < 0.85. Table 5 presents the results of the analysis QLEN instrument items using the Rasch Model.

Table 5

Analysis of Item Fit Using Rasch Model

Item	Measure	Outfit MNSQ	Outfit ZSTD	Pt Mean Corr	Item	Measure	Outfit MNSQ	Outfit ZSTD	Pt Mean Corr
SC1	**-0.66**	**1.03**	**0.4**	**0.69**	**TO1**	**-0.32**	**0.98**	**-0.3**	**0.67**
SC2	-0.37	0.99	-0.2	0.63	TO2	0.73	0.98	-0.2	0.63
SC3	0.29	0.98	-0.2	0.71	TO3	0.29	0.96	-0.6	0.68
SC4	0.76	0.93	-1.0	0.72	TO4	-0.35	0.95	-0.7	0.72
SC5	-0.08	0.93	-1.1	0.66	TO5	-0.13	0.93	-1.1	0.68
SC6	0.44	1.06	0.9	0.69	TO6	-0.61	0.89	-1.8	0.70
SC7	-0.46	0.96	-0.5	0.71	TO7	-0.34	0.98	-0.2	0.70
SC8	-1.21	0.93	-1.0	0.65	TO8	0.30	1.03	0.6	0.64
SC9	-0.07	1.02	0.3	0.68	TO9	0.11	1.04	0.8	0.64
SC10	0.05	1.02	0.3	0.68	TO10	-0.87	0.97	-0.4	0.66
TS1	0.23	1.02	0.4	0.66	AT1	-0.94	0.86	-1.7	0.69
TS2	0.09	1.15	**2.5**	0.63	AT2	0.80	0.95	-0.8	0.68
TS3	-0.16	1.07	1.2	0.65	AT3	-0.02	1.02	0.4	0.69
TS4	0.01	1.05	0.8	0.60	AT4	0.80	1.02	0.3	0.68
TS5	-0.41	1.00	0.0	0.68	AT5	0.18	0.91	-1.3	0.74
TS6	0.14	0.98	-0.4	0.65	AT6	-0.17	1.01	0.2	0.67
TS7	0.17	0.95	-0.8	0.71	AT7	-0.83	1.10	1.4	0.64
TS8	0.04	1.00	0.1	0.68	AT8	-0.73	1.04	0.6	0.59
TS9	-0.64	0.99	-0.1	0.66	AT9	-0.02	1.00	0.0	0.69

Item	Measure	Outfit MNSQ	Outfit ZSTD	Pt Mean Corr	Item	Measure	Outfit MNSQ	Outfit ZSTD	Pt Mean Corr
TS10	0.87	1.02	0.3	0.66	AT10	-0.70	1.01	0.3	0.65
EQ1	0.49	0.99	-0.1	0.68	IN1	0.62	1.02	0.4	0.67
EQ2	-0.22	1.01	0.2	0.63	IN2	0.15	1.02	0.3	0.65
EQ3	0.16	1.08	1.4	0.66	IN3	0.50	1.05	0.8	0.69
EQ4	-0.72	0.98	-0.4	0.65	IN4	0.56	0.90	-1.6	0.71
EQ5	0.75	1.00	0.0	0.65	IN5	-0.65	0.95	-0.8	0.69
EQ6	-0.68	0.88	-1.8	0.71	IN6	0.16	1.00	0.1	0.69
EQ7	0.25	0.95	-0.7	0.69	IN7	0.72	0.99	-0.2	0.68
EQ8	-0.44	0.93	-1.3	0.68	IN8	0.61	0.95	-0.8	0.66
EQ10	0.82	0.97	-0.4	0.69	IN9	0.36	1.08	1.4	0.63
					IN10	-0.32	0.98	-0.3	0.67

Of the 59 items that have been analyzed, there is one item that does not meet the criteria for item that fit the Rasch Model. According to Boone, items that do not fit should be eliminated in order to produce good instrument items and fit the model (Boone et al., 2014). There was one item eliminated from the instrument because the ZSTD outfit was >2.0. Thus, there were 58 items that meet the criteria so that they can be said to be valid items. The measure column shows the measure of item difficulty which is explained by the logit value of each item. The most difficult item is shown by item TS10 which has the highest logit value of 0.87 logit, while the easiest item is item SC8 which has the lowest logit value of -1.21 logit.

Statistical Summary

The item's difficulty level is higher than the individual's ability, as Table 6 demonstrates, and the person measure value is less than 0.07 logit than the item measure. This suggests a tendency for students' abilities to be smaller than the difficulty level of the question. Reliability for both people and things are estimated in the Rasch model. The item's and the person's respective reliability scores are 0.99 and 0.98. The quality of the items in the excellent category and the constancy of respondents' responses in the good category are both demonstrated by this reliability value. However, a very good interaction between the subject and the item was indicated by the Cronbach's Alpha score of 0.98. Viewed from person and thing, the separation index typically ranges from 0 to infinity (I. Testa, et al., 2020). Person separation shows how well the instrument can classify responses into groups. According to Testa (2020) a person separation value of less than two implies that the scale is unable to differentiate between respondents who performed well and poorly on the relevant concept. Person separation instrument of 6.75 indicating a strong capacity to separate. The QLEN's psychometric qualities appear to be very good, as indicated by its item separation of 10.67.

Table 6

Statistical Summary of QLEN

	Person	Item
N	566	59
Mean	173	1659.3
Measure	-0.07	0.00
SD	47.7	258.6
Separation	6.75	10.67
Reliability	0.98	0.99
Cronbach's Alpha	0.98	

Differential Item Functioning (DIF)

DIF is used to determine whether the QLEN scale's items are biased toward particular characteristics of respondents. If the probability value of an item in QLEN is less than 5%, it is considered to be biased (Linacre JM, 2012). DIF analysis was conducted to determine whether there were items that were biased towards gender and students who went to public or private schools. The resulting probability value shows that there are several items that have a value below 5%. In the analysis of gender bias, item AT6 "I write back the material that the teacher explains in a language that I understand" produces a probability value of 0.017 which means it is biased towards gender. On item bias analysis of school status, item AT5 "I outline the steps of solving a story problem that is quite long" of 0.0325 and item IN9 "When doing group work, my friends and I try to do it well according to the directions given by the teacher" value of 0.0360 was biased against status of school.

Discussion

This research conducted CFA and Rasch model to develop and validate psychometric properties as a result of validation an instrument that measures learning environment of numeracy. The data analysis process used CFA and Rasch model. In the testing stage, CFA is used to test the construct validity of the instrument aims to illustrate how well indicators can be used as latent variable measurement instruments through measurement models (Cimino et al., 2020). CFA also provides information on whether the model is fit to measure the numeracy learning environment measuring instrument through the goodness of fit (GOF) measure. The results of the analysis show that the instrument has met the GOF index criteria, namely $\chi 2/df = 3.14$, CFI = 0.96, AGFI = 0.86, RMSEA = 0.072 and p-value = 0.00. With the fulfillment of these criteria, it means that the data fit the measurement model. This is in line with Hermann's statement that data fit the measurement model if it meets the GOF size criteria (Herrmann et al., 2017). Furthermore, an analysis is carried out to produce a factor loading value which is a representation of the correlation between items and factors. From the analysis results, the loading factor value of all items is between 0.28 to 0.71. There is one item that is below the minimum loading factor value criteria specified, namely item EQ9, which is 0.28 and is not included in the next analysis process.

Analysis using the Rasch model began by evaluating the assumption of unidimensionality through principal component analysis (PCA) of the residuals used by raw variance explained by measures and unexplained variance at first contrast since the proposed constructs only assessed a single trait. The study's conclusion of unidimensionality just confirms that the constructs are sufficiently one dimensional for useful assessment. After the assumption of unidimensionality is met, the next stage is analyzed items that fit the Rasch model. Analysis with Rasch modeling evaluates the fit of items to the model, also known as item fit. Item fit determines whether instrument items function normally when taking measurements; if not, they should be corrected or replaced. According to Boone, fit items are items that meet the criteria of $0.5 < MNSQ < 1.5$, $-2.0 < ZSTD < +2.0$ and $0.4 <$ Pt Mean Corr < 0.85. There was one item that does not meet the criteria for item that fit the Rasch Model. To create high-quality instrument objects and match the model, this item needs to be removed. However, the Cronbach's Alpha score of 0.98 suggested a very good interaction between the item and the participant. Meanwhile, the Rasch model estimates reliability for both individuals and objects. Reliability scores for the item and the individual are 0.99 and 0.98, respectively.

For every construct, the reliability and separation indices of individuals and items were examined through summary statistics. The quality of the instrument employed, the interaction between the person and the item, and the overall response patterns provided by the respondents are all covered by summary statistics. The item separation and item reliability indices were within the recommended ranges. The person separation and dependability indices, however, fell short of the levels recommended by writers like Malec et al. (2007). Poor person separation and poor person reliability can be caused by a variety of factors, but one frequent cause is a lack of objects for a specific construct. Including items in the constructs is one way to improve the person separation and person reliability indices. This study produced strong person separation and item separation characterized by analysis results of 6.75 and 10.67. A large separation value indicates that the instrument can classify between respondents who performed well and poorly on the relevant concept.

A valid measurement is based on its instrument being unbiased. If it is found that a person with a certain attribute benefit more than another person with that attribute, the instrument or item is called biased. For example, male respondents find it easier to answer a particular instrument than female respondents; this indicates that the instrument is gender biased. Differential functioning items (DIF) are used in Rasch modeling to identify this bias. In this study, DIF is used to assess whether the items on the QLEN scale are biased towards certain traits of the respondents, namely gender and school status of the respondents, public or private. An item is considered biased if its probability value in QLEN is less than 5%. The results of the DIF analysis showed that there were three biased items, namely item AT6 "I rewrite the material explained by the teacher in a language that I understand" resulting in a probability value of 0.017 which means bias towards gender, and item AT5 "I outline the steps for solving story problems that are quite long" by 0.0325 and item IN9 "When working on group assignments, my friends and I try to do well according to the directions given by the teacher" by 0.0360 which means bias towards school status. These items need to be revised again so that the resulting instrument becomes more perfect.

Conclusions

Studies have been done on the development and validation of QLEN instrument. 58 of 60 items were valid item and had the ability to measure a single dimension and exhibit good overall performance. Three of them showed a bias in favor of the characteristics of the respondent. These three items require writing improvements to produce an instrument that is free from item bias. Analyzing of measure showed that the item's difficulty level is higher than the individual's ability. It indicates a tendency for student ability to be lower than the difficulty level of the question. The reliability of the instrument is in the excellent category both in the item and person reliability categories. Nevertheless, assessments utilizing the CFA and Rasch models demonstrate that the QLEN instrument has sufficient psychometric qualities to assess numeracy learning environment. The QLEN instrument has strong psychometric qualities according to both the Rasch model and CFA, as demonstrated by this study.

Acknowledgmets

We would like to express our gratitude to the school for giving the researchers the opportunity to conduct research at the school. In addition, the researchers do not forget to give high appreciation to all parties who have provided assistance in various forms in the implementation of research so that the writing of this article can be completed properly.

References

Aldridge, J. M., Fraser, B. J., Bell, L., & Dorman, J. (2012). Using a new learning environment. Questionnaire for reflection in teacher action research. *Journal of Science Teacher Education*, *23*(3), 259–290. https://doi.org/10.1007/s10972-012-9268-1

Aluja-Banet, T., Sancho, M. R., & Vukic, I. (2019). Measuring motivation from the virtual learning environment in secondary education. *Journal of Computational Science*, *36*. https://doi.org/10.1016/j.jocs.2017.03.007

Aluri, V. L. N., & Fraser, B. J. (2019). Students' perceptions of mathematics classroom learning environments: measurement and associations with achievement. *Learning Environments Research*, *22*(3), 409–426. https://doi.org/10.1007/s10984-019-09282-1

Andrade, C. (2020). Sample size and its importance in research. *Indian Journal of Psychological Medicine*, *42*(1), 102–103. https://doi.org/10.4103/IJPSYM.IJPSYM_504_19

Ayu, S. M., Gustina, E., Sofiana, L., Wardani, Y., & Sukarelawan, M. I. (2023). Physical and psychological violence victimization scale in adolescents dating: Confirmatory factor analysis and Rasch model. *International Journal of Evaluation and Research in Education*, *12*(1), 96–105. https://doi.org/10.11591/ijere.v12i1.22250

Bisri, H., Rahmawati, Y., & Ridwan, A. (2023). Inventory for measuring higher-order thinking learning environment. *New Educational Review*, *73*, 162–174. https://doi.org/10.15804/tner.23.73.3.12

Boone, W. J., & Staver, J. R. (2020). Advances in Rasch analyses in the human sciences. In J. Staver & W. J. Boone (Eds.), *Advances in Rasch analyses in the human sciences* (pp. C1–C2) https://doi.org/10.1007/978-3-030-43420-5_21

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences.* Springer. http://dx.doi.org/10.1007/978-94-007-6857-4

Buljan, I., Tokalić, R., Marušić, M., & Marušić, A. (2019). Health numeracy skills of medical students:Cross-sectional and controlled before-and-after study. *BMC Medical Education*, *19*(1), 1–10. https://doi.org/10.1186/s12909-019-1902-6

Cahyani, C. D., & Mohammad, A. (2023). Systematic literature review: Application of probing prompting learning model to mprove students' numeracy skill. *Jurnal Ilmiah Soulmath : Jurnal Edukasi Pendidikan Matematika*, *11*(1), 1–12. https://doi.org/10.25139/smj.v11i1.4596

Çelik, H. C., & Özdemir, F. (2020). Mathematical thinking as a predictor of critical thinking dispositions of pre-service mathematics teachers. *International Journal of Progressive Education*, *16*(4), 8–98. https://doi.org/10.29329/ijpe.2020.268.6

Cimino, A. N., Killian, M. O., Von Ende, A. K., & Segal, E. A. (2020). Measurement models in social work research: A data-based illustration of four confirmatory factor models and their conceptual application. *British Journal of Social Work*, *50*(1), 282–301. https://doi.org/10.1093/bjsw/bcz164

Davis, D. R., & Boone, W. (2021). Using Rasch analysis to evaluate the psychometric functioning of the other-directed, lighthearted, intellectual, and whimsical (OLIW) Adult playfulness scale. *International Journal of Educational Research Open*, *2*(January), 100054. https://doi.org/10.1016/j.ijedro.2021.100054

Garcia-Retamero, R., Sobkow, A., Petrova, D., Garrido, D., & Traczyk, J. (2019). Numeracy and risk literacy: What have we learned so far? *Spanish Journal of Psychology*, 1–11. https://doi.org/10.1017/sjp.2019.16

Hair, J., Anderson, R., Babin, B., & Black, W. (2010). *Multivariate data analysis* (7th ed.). Cengage.

Herrmann, K. J., Bager-Elsborg, A., & Parpala, A. (2017). Measuring perceptions of the learning environment and approaches to learning: Validation of the learn questionnaire. *Scandinavian Journal of Educational Research*, *61*(5), 526–539. https://doi.org/10.1080/00313831.2016.1172497

Ilias, A., & Siew, N. M. (2022). Development of future thinking test instrument for secondary school students. *International Journal of Education, Psychology and Counseling*, *7*(47), 337–353. https://doi.org/10.35631/IJEPC.747029

Insorio, A. O., & Librada, A. R. P. (2020). Enhancing mathematical critical thinking and problem-solving skills through emergenetics as a grouping mechanism. *Contemporary Mathematics and Science Education*, *2*(1), 1–10. https://doi.org/10.30935/conmaths/9289

Koul, R. B. (2023). Classroom learning environments and assessment practices in science classrooms in western Australia. In R. Maulana, M. Helms-Lorenz, Robert M. Klassen (Eds.), *Effective teaching around the world: theoretical, empirical, methodological and practical insights* (pp. 317–337). https://doi.org/10.1007/978-3-031-31678-4_15

Linacre JM. (2012). *Winsteps Rasch measurement computer program user's guide*. https://www.winsteps.com/winman/copyright.htm

Md Hassan, N., Majid, N. A., & Hassan, N. K. A. (2020). Validation of learning environment inventory for secondary school contexts. *International Journal of Evaluation and Research in Education*, *9*(2), 379–384. https://doi.org/10.11591/ijere.v9i2.20444

Memon, M. A., Ting, H., Cheah, J.-H., Thurasamy, R., Chuah, F., & Cham, T. H. (2020). Sample size for survey research: review and recommendations. *Journal of Applied Structural Equation Modeling*, *4*(2), 1–20. https://doi.org/10.47263/JASEM.4(2)01

Mustafa, N., Khairani, A. Z., & Ishak, N. A. (2021). Calibration of the science process skills among Malaysian elementary students: A Rasch model analysis. *International Journal of Evaluation and Research in Education*, *10*(4), 1344–1351. https://doi.org/10.11591/ijere.v10i4.21430

Omar, M. S., & Awang, M. I. (2023). The learning environment as a predictor of higher order thinking skills. *International Journal of Evaluation and Research in Education*, *12*(1), 395–402. https://doi.org/10.11591/ijere.v12i1.23959

Pergo, A. T. C., França, A. B., Teixeira, P. H. M., Siqueira, L. B., Cortez, P. J. O., & Vitorino, L. M. (2023). Development and validation of a vaccine hesitancy scale for COVID-19 to the Portuguese language. *Research, Society and Development*, *12*(3), 1–9. https://doi.org/10.33448/rsd-v12i3.40357

Rusticus, S. A., Pashootan, T., & Mah, A. (2023). What are the key elements of a positive learning environment? Perspectives from students and faculty. *Learning Environments Research*, *26*(1), 161–175. https://doi.org/10.1007/s10984-022-09410-4

Sasha Nikolic, Montserrat Ros, K. J. & Z. S. (2021). Remote, simulation or traditional rngineering teaching laboratory: A systematic literature review of assessment implementations to measure student achievement or learning. *European Journal of Engineering Education*, *46*(6), 1141–1162. https://doi.org/10.1080/03043797.2021.1990864

Solé-Beteta, X., Navarro, J., Gajšek, B., Guadagni, A., & Zaballos, A. (2022). A Data-driven approach to quantify and measure students' engagement in synchronous virtual learning environments. *Sensors*, *22*(9), 1–27. https://doi.org/10.3390/s22093294

Testa, I., Capasso, G., Colantonio, A., Galano, S., Marzoli, I., Uccio, U. S., & Serroni, G. (2020). Validation of university entrance tests through Rasch analysis. In Khine M.S. (Ed.), *Rasch measurement:applications in ouantitative educational research* (pp. 99–124). https://doi.org/10.1007/978-981-15-1800-3_3

Skaičiavimo mokymosi aplinka: instrumento kūrimas ir jo validumas pagal patvirtinančiąją faktorinę analizę ir Rasch'o modelį

Cynthia Mayang Sari[1], Achmad Ridwan, Ridwan[2], Yuliatri Sastrawijaya, Sastrawijaya[3]

[1] Valstybinis Džakartos universitetas, Švietimo tyrimų ir vertinimo katedra, Jl., Rawamangun Muka g. 11, Ravamangunas, Pulo Gadungas, ID-13220 Rytų Džakarta, Indonezija, cynthia.mayang20@gmail.com

[2] Valstybinis Džakartos universitetas, Švietimo tyrimų ir vertinimo katedra, Chemijos mokymo katedra, Jl., Rawamangun Muka g. 11, Ravamangunas, Pulo Gadungas, ID-13220 Rytų Džakarta, Indonezija, achmadridwan@unj.ac.id

[3] Valstybinis Džakartos universitetas, Švietimo tyrimų ir vertinimo departamentas, Informatikos ir kompiuterių inžinerijos mokymo katedra, Jl., Rawamangun Muka g. 11, Ravamangunas, Pulo Gadungas, ID-13220, Rytų Džakarta, Indonezija, yuliatri@unj.ac.id

Santrauka

Šio tyrimo tikslas – sukurti ir išbandyti tyrimo instrumentą, klausimyną (angl. *QLEN*), kad būtų galima įvertinti skaičiavimo mokymosi aplinką. Bandomajame tyrime dalyvavo 264 studentai, o galutiniame tyrime dalyvavo 566 studentai. Tyrimo instrumentui įvertinti buvo naudojama patvirtinančioji faktorinė analizė (angl.*CFA*) ir Rasch'o modelis. Tyrimas parodė, kad 58 iš 60 QLEN elementų buvo pagrįsti ir atitiko Rasch'o modelį, elementų ir asmens patikimumo vertės buvo atitinkamai aukštos – 0,99 ir 0,98. QLEN skalė parodė stiprias psichometrines savybes ir tinkamą faktorių struktūrą. Kurti QLEN instrumentą paskatino skaičiavimo mokymosi aplinkos svarba formuojant mokinių skaičiavimo įgūdžius. Būtent ankstesni tyrimai rodė, kad trūksta supratimo ir skaičiavimo mokymosi aplinkos, todėl šiam tikslui pasiekti pasirinkta sukurti tinkamą ir patikimą instrumentą. QLEN buvo sukurtas remiantis šešiais matmenimis iš 60 elementų; visų pagrįstumas ir patikimumas buvo įvertintas. Tyrime buvo naudojamas patvirtinančiosios faktorinės analizės ir Rasch'o modelio derinys, siekiant įvertinti QLEN instrumento psichometrines savybes. Patvirtinančiosios faktorinės analizės rezultatai parodė gerą atitikimą QLEN faktoriaus struktūrai, o įvairūs tinkamumo indeksai atitiko kriterijus. Rasch'o modelio analizė atskleidė, kad iš 60 elementų 58 buvo geri ir tinkami modeliui. Vis dėlto skirtingų elementų veikimo (angl. *DIF*) analizės metu buvo nustatyti trys šališki elementai, vadinasi, norint išvengti šališkumo, reikia šiuos elementus peržiūrėti.

Esminiai žodžiai: *skaičiavimo mokymosi aplinka, Rasch'o modelis, patvirtinančioji faktorinė analizė.*

Gauta 2024 02 01 / Received 01 02 2024
Priimta 2024 04 02 / Accepted 02 04 2024