

Ethnoscience-Quizizz test to measure problem-solving skills: a Rasch analysis

Khoirun Nisa, Nadi Suprpto, Mohd Zaidi Amiruddin, Eka Putri Dian Nata Sari, Bilqisth Dwi Athiah

Department of Physics Education, Faculty of Mathematics and Natural Science, Universitas Negeri Surabaya, Surabaya, Indonesia

Article Info

Article history:

Received Jul 15, 2023

Revised Oct 5, 2023

Accepted Mar 18, 2024

Keywords:

Ethnoscience
Problem-solving
Quizizz
Rasch
Test

ABSTRACT

Problem-solving skills (PSS) are essential for students to face challenges in the 21st century. Based on PISA results, the PSS of Indonesian students are in a low category. Not only that, a few assessments of PSS have been developed. This study aims to analyze the quality of the ethnoscience-Quizizz test (EQT). The research method used is R&D with the ADDIE model. The data analysis technique was based on the Rasch analysis using the Ministep software. The item and person reliability are 0.9 (good) and 0.40. Cronbach's alpha has a value of 0.53. The MNSQ-ZSTD infit-outfit score is in a suitable category. Each item is relatively accurate and do not have DIF. In addition, the EQT is suitable to measure students' PSS because it has passed the criteria. The EQT incorporates indicators for problem-solving, making it appropriate for 21st-century learning with an ethnoscience "Kerapan Sapi" problem context. Future studies can use this problem-solving test to determine the profiles of students' problem-solving abilities using Rasch analysis.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Nadi Suprpto

Department of Physics Education, Faculty of Mathematics and Natural Science, Universitas Negeri Surabaya

St. Ketintang, Gayungan, Surabaya-60231, Indonesia

Email: nadisuprpto@unesa.ac.id

1. INTRODUCTION

The Indonesian Merdeka curriculum was concerned with Pancasila's characters that integrated with culture. Meanwhile, local wisdom is knowledge, ideas, culture, and insights inherited and maintained as an identity and guidance for human life [1]. Ethnoscience is learning strategies that combine local wisdom and scientific knowledge [2], [3]. Ethnoscience learning can create a new experience for students to increase self-competence. Specifically, physics is subject to life-physics problems or phenomena, for example, ethnophysics [3]. *Kerapan Sapi* is a traditional bull racing festival with potency and interrelationships with science material [4]. Ethnoscience integrated local wisdom has the potential to enhance students' problem-solving skills (PSS). PSS is the ability of trouble-fixing steps to understand problems, develop strategies, and examine the solution [5]. Teachers teach their students to think and practice solving problems so students can improve their emotional, cognitive, and psychomotor. There are five PSS indicators: problem assistance, design a concept, conceptualize the strategy, implement the solution, and examine the solution [6], [7]. In the 21st century, students need PSS to solve their problems. So, teacher must conduct instrument test-based problems to measure students' PSS and evaluate the learning process [8]–[10].

The instrument test design must incorporate the conceptual assessment framework (CAF), composed of models for competency, evidence, tasks, and assembly. The CAF comprise the entire evaluation [11], [12]. Proper evaluation procedures lead to accurate measurement outcomes [13]. So, the validity is one of the requirements that the instrument test must fulfill [14]. Therefore, to eliminate measurement mistakes

and assess the actual state of the students, teachers need reliable exam tools [15]. Indonesian students have a low PSS rate, especially in physics subjects [16]. There currently needs to be more instruments to measure cognitive levels C4 through C6 [17]. However, the questions on PSS techniques fall between cognitive levels C4 and C6.

On the other hand, globalization significantly impacted innovation technology in education [18], [19]. Game-based learning especially Quizizz is an innovative method of delivering the material or test and improve students' learning results [20]–[23]. According to Figure 1, the authors analysis research trend with science problem-solving assessment on work and energy material based on the Scopus database for the last five years. The data was recorded on October 2022. There are seven clusters such as the physics material used in assessment, primarily work and energy material (red, blue, and yellow colors), assessment and learning (purple and yellow colors), the subject research of assessment (green color), and the research method of assessment (blue color).

Figure 1 describes many researchers still focus on developing assessment-based PSS [24], [25]. In addition, development research of instruments test of local wisdom has been carried out. However, research on work and energy (physics concept) integrated ethnoscience has yet to be conducted so far [26], [27]. This paper aims to present the development and analyze the quality of an instrument test. The ethnoscience test in work and energy material, EQT, must be able to characterize PSS' students in order to as a test tool. An analysis of item bias in EQT was conducted using attributes from the Rasch model [28]. The measurement findings will be influenced and not accurately reflect the situation if this happens. A phenomenon known as bias occurs when a group's characteristics have an impact on test outcomes [29]. Additionally, bias is a sign of unfairness, inconsistent results, and contaminated test results [30]. Revisions to items are based on the results of Rasch's analysis, including item fit statistics, item category structure, differential item functioning (DIF), person-item maps (wright maps), and dimensionality. The fit of the items with the model is determined by looking at the mean square residual (MNSQ) and standardized mean square residual (ZSTD).

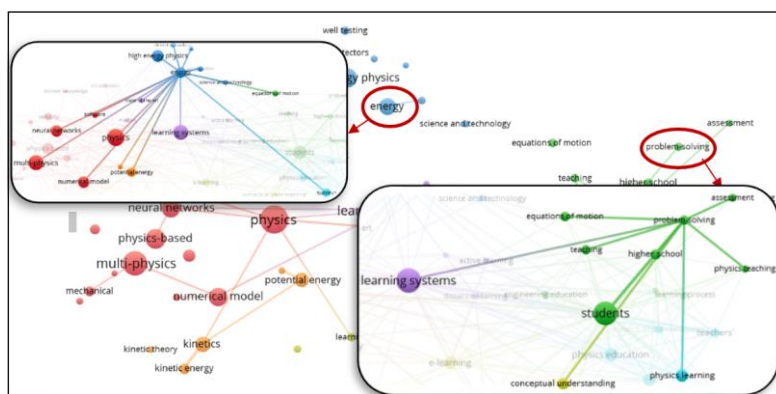


Figure 1. Research trend of physics problem-solving test

2. METHOD

The type of research is development research using the ADDIE model, as presented in Figure 2, which consists of designing, developing, and testing the validation product [31]. This paper conducts a problem-solving instrument test based on ethnoscience Madura's Bull Racing and analysis its validity using Rasch Analysis. There are 35 second-year of senior high school students. This research chooses participants using purposive sampling. Participants were 28% male students and 72% female students.

Analyze stage concludes with material analysis and curriculum or indicator of learning analysis. After selecting the indicator, researchers designed the ten problem-solving tests integrated with ethnoscience and Quizizz as assessment media. The indicators of students' problem-solving skills are defining the problem, checking the problem, planning the solutions as shown in Table 1. Category of unidimensionality value, implementing the problem, and evaluating the solutions [32]. The instruments contain Bloom's taxonomy from C3 (apply) to C5 (evaluate) related to problem-solving indicators. At the development stage, researchers searched for information and material about Madura's culture and related physics concepts, making illustrations, graphics, and charts, then inserting them into Quizizz. In the development stage, researchers conduct the assessment instruments that contain an introduction, guidelines, question items, answer choices, answer sheets, answer keys, and scoring guidelines. At this stage, the assessment product was analyzed using Ministep software (Rasch analysis) [33]–[35]. The Ministep operation results include dimensionality, item fit, reliability, category fit, Wright map, and differential item functioning (DIF) [36]. It represents those that will provide information about the quality of the respondents, instruments, and interactions between persons

and items [37]. The last stage is evaluations; researchers provide the test results based on Rasch analysis output. Values for validity and reliability are derived from student test results. The validity value is calculated by applying the unidimensionality value to the empirical raw variance explained by measures. Table 1 shows the unidimensionality value requirements.

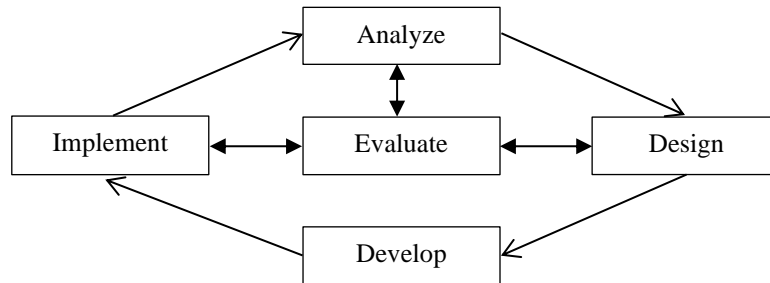


Figure 2. ADDIE model [38]

The reliability value of the item, the person, and Cronbach’s alpha determine the reliability value. The value of a person’s ability and item reveals the consistency of respondents’ responses and the caliber of the items [39]. In contrast, the value of Cronbach’s alpha shows the interaction between the person and item as a whole and measures the reliability value [40], [41]. The specifications for the person-item reliability and Cronbach alpha (CA) are listed in Table 2.

Further analysis includes the Wright map, item measure, DIF, and item fit order: i) Wright map: a logit scale of items and persons is shown on the wright map. The left half of the map displays the distribution of student ability levels, which is a depiction of students' PSS talents, while the right side of the map displays the distribution of item difficulties; ii) The item measure displays the questions' level of difficulty as stated on a logit scale [39]; iii) Item fit order to determine the item fits or misfits [42]. The item must at least achieve MNSQ (outfit mean square) value: $0.5 < MNSQ < 1.5$ and ZSTD (outfit Z-standard) value: $-2.0 < ZSTD < +2.0$ [43]–[46]. The item misfit will be removed or repaired [26]; and iv) $DIF < 0.05$ for item fit.

Table 1. Category of unidimensionality value

Value of RVEM	Description
$x < 20\%$	Invalid
$20\% < x \leq 40\%$	Enough
$40\% < x \leq 60\%$	Valid
$x > 60\%$	Excellent

Note: RVEM= Raw variance explained by measures

Table 2. Category of item-person reliability and CA

Value of reliability	Description	Value of CA	Description
$x < 0.67$	Weak	$x < 0.50$	Weak
$0.67 < x \leq 0.80$	Enough	$0.50 < x \leq 0.60$	Enough
$0.80 < x \leq 0.90$	Good	$0.60 < x \leq 0.70$	Good
$0.90 < x \leq 0.94$	Very good	$0.70 < x \leq 0.80$	Very good
$x > 0.94$	Excellent	$x > 0.80$	Excellent

3. RESULTS AND DISCUSSION

3.1. Design of ethnoscience-Quizizz test (EQT)

Kerapan Sapi is a Madurisch community event every year in August or September. This race is known as two pairs, the bull that each jockey mounts is raced fast (*ê-kerrap*) with the usage of equipment (*kalêlêlês*). The cow is forced to run on a track approximately 200 meters which can be completed in 14 to 18 seconds. When the bull race and the jockey have a small mass, the bull can race fast. It means the bull's acceleration is significant. According to Newton’s 2nd law, force is proportional to mass and acceleration. The force comes from the *trekeng*'s pressure and the leg muscles of the bull. The direction of force and acceleration are the same. Thus, the friction force between *Kaleles* and runway ground. In addition, bulls have kinetic energy, which has a mass and velocity if it is a race. When bull race, there is a working concept where an activity carried out by a bull uses a certain amount of energy. This concept exists when the bull race competes. The purpose of planning the instrument is to ensure integrity between physics (work and energy)

concepts, ethnoscience, and Bull Racing as Madura’s culture. Work and energy materials were chosen because they integrated with Madura’s Bull Racing. The results of the materials and competencies analysis are presented in Table 3. The results of the study on the correlation between work-energy concept and ethnoscience are presented in Table 4.

The test presented problems in Madura's Bull Racing, consisting of 10 items and four answer choices for each item. The assessment system using Quizizz was developed as a random question-and-answer feature, so each student would get a different sequence of question numbers and answer choices. When students work on questions and choose answers, a red indicator will appear for the wrong student's answer and a green indicator for the correct answer. Every time students finish working on a question, a ranking of student acquisitions will appear. After students work on all the questions, they can review the results and find the correct answers to questions that were answered incorrectly by students. Then, the scores obtained by students accumulated automatically. The examples of the test are indicated in Figure 3.

Table 3. Analysis of material and competencies

Materials	Competencies	The study of Madura’s bull racing
Work and energy	Examine the ideas of energy, work, and energy relationships, the law of energy conservation, and how it applies to commonplace occurrences. Offer solutions to motion problems that arise in daily life by using the scientific method.	Energy and work The law of conservation energy

Table 4. Concept of EQT

Indicator	Cognitive domain	Stimulus	The study of physics concept	The study of Madura’s bull racing
Evaluate the minimum energy required for cow	C4	Data bull racing’s participants	Work	Acceleration of bull is equal to its work
Evaluate and create a physics equation on Madura’s Bull Racing	C4	Data on herbal drink	Work and calories	The herbal drink can transfer calories to cow
Evaluate work based on the direction of the force and displacement	C5	Bull racing events	Work	Bull run in the field with force and distance
Predict the state of an object that has maximum kinetic energy	C5	Tracking event	Kinetic energy	Bull run with velocity in track
Analyze the kinetic energy graph	C4	Story of bull run	Kinetic energy	Bull run with velocity in track
Analyze potential energy and its change at each point	C5	Car transport equipment	Potential energy	Car goes through a sub-district with some high
Designed the solutions for minimum mechanical energy	C5	Car transport equipment	Mechanical energy	Car goes through a sub-district with velocity
Analyze the correlation between work and energy	C3	Tracking event	Correlation work and kinetic energy	Bull run with velocity in track
Analyze the comparison of the value of work between two objects	C4	Two bulls run into the field	Comparison of work between 2 objects	Bull run with different times in track

1. The winner of the bull race is a pair of cows who can cross the finish line first within 150 m. A researcher observed the competition and obtained the following data in the table below. The cow that has the minimum work is ...

Team	m _{cow} (kg)		m _{jockey} (kg)	m _{kaleles} (kg)	a _{cow} (m/s ²)
	1	2			
<i>Jetmatic</i>	250	255	25	2	1.5
<i>Bandar Jawa</i>	256	255	23	2	1
<i>Baby Selor</i>	250	252	22	2	2
<i>Tappor Kelap</i>	254	253	24	2	1.3

a. *Jetmatic* team b. *Bandar Jawa* team c. *Baby Selor* team d. *Tappor Kelap* team

2. The distance between the start and finish lines is 150 m. Two pairs of bulls will compete in the first stage. The Trobos Sakti team is seen running on a line straight to the finish line with a force of 1120 N. While the Baby Selor team is seen running in the opposite direction, The path of the Milky Trobos Cow starts with an angle between the direction force and its track is 60°, then moves in a straight line after the Milky Trobos Cow takes 3/5 tracks with a constant force of 1125 N. The minimum work will occur that...

a. Trobos Saktis' team ran from start to finish. c. After the Baby Selors' team overtook the Trobos Sakti team.
 b. Baby Selors' team will overtake Trobos Sakti team. d. Baby Selors' team ran from start to finish.

Figure 3. Example of EQT

3.2. Analysis of EQT as an instrument test

3.2.1. Validity and reliability of EQT

The result of validity is based on unidimensionality, as seen in Figure 4. The dimensional test is intended to detect other dimensions in the item. Raw variance is above 40%, meaning the distribution of item

difficulty levels is good. Based on the results of the Rasch analysis, a data variance of 45.7% was found, which indicated a good distribution of item difficulty levels. Eigenvalue 3.0 indicates that the goods are either good or not mixed with other dimensions. So, the test instrument developed is very valid and suitable for physics learning, especially in work and energy. The compatibility between the material, media, and ethnophysics given is good. The physics learning assessment instrument using Quizizz is feasible.

Reliability represents the level of consistency of the assessment test in measuring data repeatedly, and it will give relatively the same results. Figure 5(a) shows person reliability is 0.40, indicating a lack of consistency in the student's answers. The correlation between the items and the subject is shown by Cronbach's alpha, which is 0.53 and is considered good. While, Figure 5(b) indicates the reliability of the item value is 0.90, indicating that *Kerapan Sapi* Madura's integrated problem-based question items fall under the very good category. The item's mean difficulty was lower than the person's, and the subject's average measure was 1.17 logit (which was set at 0.00 logit). Students can take the test more easily as a result.

	Eigenvalue	Observed	Expected
Total raw variance in observations =	18.4122	100.0%	100.0%
Raw variance explained by measures =	8.4122	45.7%	44.3%
Raw variance explained by persons =	3.2480	17.6%	17.1%
Raw Variance explained by items =	5.1642	28.0%	27.2%
Raw unexplained variance (total) =	10.0000	54.3%	55.7%
Unexplned variance in 1st contrast =	2.4760	13.4%	24.8%
Unexplned variance in 2nd contrast =	1.9785	10.7%	19.8%
Unexplned variance in 3rd contrast =	1.4915	8.1%	14.9%
Unexplned variance in 4th contrast =	1.1461	6.2%	11.5%
Unexplned variance in 5th contrast =	1.0665	5.8%	10.7%

Figure 4. Unidimensionality of EQT

SUMMARY OF 35 MEASURED (NON-EXTREME) Person									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	6.7	10.0	1.17	.96	.89	-.15	1.03	.13	
SEM	.3	.0	.23	.04	.10	.21	.24	.17	
P.SD	1.7	.0	1.37	.22	.61	1.23	1.42	.98	
S.SD	1.7	.0	1.39	.22	.62	1.24	1.44	.99	
MAX.	9.0	10.0	3.46	1.39	2.37	3.10	5.65	2.96	
MIN.	3.0	10.0	-1.39	.77	.24	-1.64	.08	-.73	
REAL RMSE	1.06	TRUE SD	.87	SEPARATION	.82	Person RELIABILITY	.40		
MODEL RMSE	.98	TRUE SD	.96	SEPARATION	.98	Person RELIABILITY	.49		
S.E. OF Person MEAN = .23									

(a)

SUMMARY OF 10 MEASURED (NON-EXTREME) Item									
	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	24.5	36.0	.00	.52	1.03	-.06	1.03	-.08	
SEM	2.9	.0	.62	.05	.08	.33	.23	.26	
P.SD	8.8	.0	1.85	.14	.25	.99	.68	.77	
S.SD	9.3	.0	1.95	.14	.26	1.04	.71	.81	
MAX.	34.0	36.0	4.24	.76	1.43	1.32	2.90	1.50	
MIN.	4.0	36.0	-2.32	.40	.67	-1.94	.53	-1.24	
REAL RMSE	.58	TRUE SD	1.76	SEPARATION	3.03	Item RELIABILITY	.90		
MODEL RMSE	.54	TRUE SD	1.77	SEPARATION	3.27	Item RELIABILITY	.91		
S.E. OF Item MEAN = .62									

(b)

Figure 5. Reliability analysis result of (a) Person and (b) Item

3.2.2. Item fit and DIF of EQT

The MNSQ-ZSTD infit-outfit score is in a good category, as shown in Table 5. It shows that the test instrument used to measure students' problem-solving skills in the context of *Kerapan Sapi* Madura is appropriate. Likewise, the value of the ZSTD outfit is included in the good category and shows that the data obtained is rational. The MNSQ and ZSTD outfit and infit values obtained also show that the distribution of items is fit and by the provisions already mentioned. It shows that the items are under the Rasch model and can be used as test instruments.

Table 5. Item fit

Item	Measure	Infit		Outfit		Ptmeasural	
		MNSQ	ZSTD	MNSQ	ZSTD	Corr.	Exp.
1	-2.32	1.06	0.29	0.55	0.09	0.23	0.23
2	-0.48	1.10	0.49	0.78	-0.23	0.40	0.42
3	-1.18	1.43	1.32	1.35	0.66	0.11	0.35
4	-2.32	1.14	0.42	1.22	0.61	0.13	0.23
5	1.53	0.91	-0.45	0.79	-0.68	0.61	0.55
6	-0.69	0.71	-1.19	0.54	-0.68	0.55	0.40
7	1.05	1.06	0.41	0.92	-0.20	0.51	0.53
8	0.08	0.67	-1.94	0.53	-1.24	0.65	0.47
9	0.08	0.82	-0.96	0.72	-0.60	0.57	0.47
10	4.24	1.41	1.01	2.90	1.50	0.18	0.48
Mean	0.00	1.03	-0.06	1.03	-0.08		
P.SD	1.85	0.25	0.99	0.68	0.77		

Items are said to have a bias (containing DIF) if the response pattern is from the group gender that shows the difference, as seen in Figure 6. If the probability P is 5% (0.05), the item contains DIF [44], [46]. Some items in Figure 6 do not have a bias (DIF) that present that EQT have good quality [47]–[49]. The EQT is an excellent fit for 21st-century learning in the context of issues that will help students become better problem solvers. The creation of test devices is a fascinating field that can inspire students to pursue physics as a subject. Because Quizizz is a gamification evaluation that incorporates avatars, memes, music, and other entertainment, it is the appropriate medium that is simple to use, practical, and fascinating [50]. Assessment using the Quizizz application is interesting because of its game concept [51]. The use of ethno-physics context and evaluation media in work and energy material during physics learning is a choice that needs to be taken seriously because its abstract [52]–[56].

Person CLASSES	SUMMARY DIF		PROB.	BETWEEN-CLASS/GROUP		Item	
	CHI-SQUARED	D.F.		UNWTD MNSQ	ZSTD	Number	Name
2	2.7612	1	.0966	4.0112	1.72	1	1
2	1.9494	1	.1627	2.1108	1.07	2	2
2	.0013	1	.9715	.0016	-1.40	3	3
2	2.7612	1	.0966	4.0112	1.72	4	4
2	.8718	1	.3505	.9220	.41	5	5
2	.6760	1	.4110	.7110	.24	6	6
2	.0781	1	.7798	.0793	-.74	7	7
2	.6021	1	.4378	.6305	.17	8	8
2	.6021	1	.4378	.6305	.17	9	9
2	.0030	1	.9567	.0028	-1.35	10	10

Figure 6. DIF item

3.2.3. Wright map

The wright map, as shown in Figure 7, maps the items' logit scale representing the item's difficulty. Three questions are categorized as complex for students to work on because they have a logit scale value above 0. Based on the variable map, questions 5, 7, and 10 are difficult for students to work on, as are questions C4 and C5, namely the level of analysis and synthesis questions.

One of the advantages of developing an instrument using the Rasch model is that it can be used more widely by other parties because of the independent nature of people and items. In addition, item analysis measures the level of item accuracy or standard error measurement (SEM). If the SEM value is 0.5 logit, then the item is accurate in measuring the SEM, and if the SEM is greater than 1.0 logit, the item is not accurate in measuring the SEM. All items have a good level of accuracy. The item indicated as a complicated item is 10, and the effortless item is 8. Wright's map still needs to include some gaps between items. To improve the quality of the instrument, if the gap is greater than 0.5 logit, the test developer can fill the gap by adding new items that are more difficult than the items below [57].

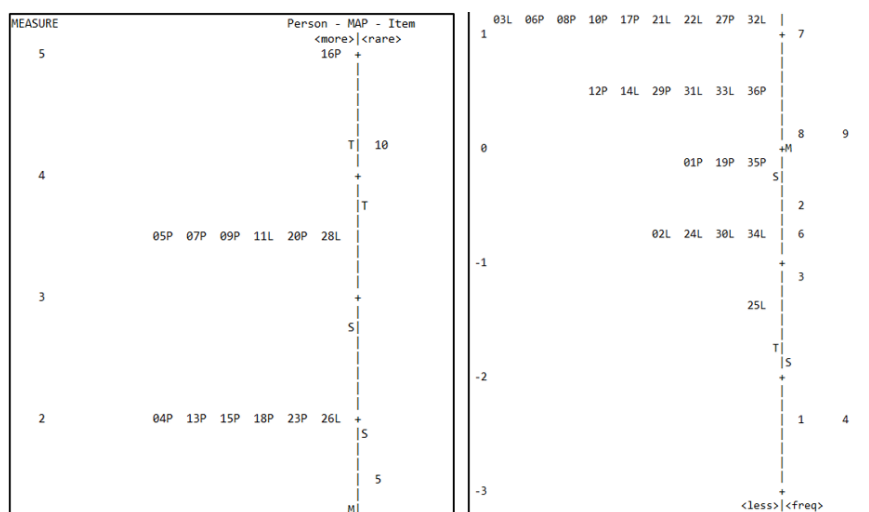


Figure 7. Wright map

4. CONCLUSION

Quizizz is an e-learning platform used for game-based learning. EQT has a high item reliability of 0.9, indicating good quality items. Then, person reliability was 0.40 and CA was 0.53. The MNSQ-ZSTD infit-outfit score is in a good category, with accurate items and have not DIF. The EQT used to measure students' PSS is appropriate for physics learning, especially work and energy material. However, limitations include not exploring respondents' regional, gender, and cultural characteristics. So, it is impossible to generalize this result to the Indonesian condition. Heterogeneity of students was required for an in-depth evaluation of EQT's bias. Future research should consider the diversity of participants, including ethnicity or culture, and explore multiple local wisdoms and subjects. Some recommendations are addressed for the next researchers: i) Rasch analysis can be used as an alternative for researchers who want to analyze items and person simultaneously; ii) Further researchers should address the person or respondents in term of region, gender, and cultural characteristics.

REFERENCES




- [1] S. N. Muhajir, S. Utari, and I. R. Suwama, "How to develop test for measure critical and creative thinking skills of the 21st century skills in POPBL?" *Journal of Physics: Conference Series*, vol. 1157, no. 3, p. 032051, Feb. 2019, doi: 10.1088/1742-6596/1157/3/032051.
- [2] M. P. Sudarmin, *Character education, ethnoscience and local waters*. Semarang: Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Semarang (in Indonesian), 2014.
- [3] N. I. Wulansari and S. Admoko, "Identification of Physics Concepts in Reog Ponorogo's Dhadak Merak Dance as A Source of Learning Physics: An Analytical Study," *Berkala Ilmiah Pendidikan Fisika*, vol. 9, no. 1, pp. 105–113, 2021, doi: 10.20527/bipf.v9i1.9862.
- [4] N. Suprpto, "Demographic sources as a local wisdom: Potency of Indonesian physics education researchers in conducting survey research," *Journal of Physics: Conference Series*, vol. 1171, no. 1, p. 012003, Feb. 2019, doi: 10.1088/1742-6596/1171/1/012003.
- [5] H. Azninda and W. Setyarsih, "Application of problem based learning model using self regulated strategy learning strategy to improve students' problem solving ability," (in Indonesian), *Inovasi Pendidikan Fisika*, vol. 7, no. 2, pp. 347–352, 2018.
- [6] C. Riantoni, L. Yulianti, N. Mufti, and N. Nehru, "Problem solving approach in electrical energy and power on students as physics teacher candidates," *Jurnal Pendidikan IPA Indonesia*, vol. 6, no. 1, pp. 55–62, 2017, doi: 10.15294/jpii.v6i1.8293.
- [7] B. K. Prahani *et al.*, "ORNE Learning Model to Improve Problem-Solving Skills of Physics Bachelor Candidates: An Alternative Learning in the Covid-19 Pandemic," *Jurnal Penelitian Fisika dan Aplikasinya (JPFA)*, vol. 10, no. 1, p. 71, Sep. 2020, doi: 10.26740/jpfa.v10n1.p71-80.
- [8] D. David, B. Kartowagiran, and S. P. Harjo, "Evaluation and Development Strategy of Senior High School Indonesisch Nerderlandsche School (Ins) Kayutanam," (in Indonesian), *Jurnal Penelitian dan Evaluasi Pendidikan*, vol. 20, no. 1, pp. 27–44, Jun. 2016, doi: 10.21831/pep.v20i1.7518.
- [9] I. Maryani, Z. K. Prasetyo, I. Wilujeng, S. Purwanti, and M. Fitriawanati, "HOTs Multiple Choice and Essay Questions: A Validated Instrument to Measure Higher-order Thinking Skills of Prospective Teachers," *Journal of Turkish Science Education*, vol. 18, no. 4, pp. 674–690, 2021, doi: 10.36681/tused.2021.97.
- [10] S. Sugiyanta and S. Soenarto, "An evaluation model of educational quality assurance at junior high schools," *REID (Research and Evaluation in Education)*, vol. 2, no. 2, pp. 194–205, Dec. 2016, doi: 10.21831/reid.v2i2.11118.
- [11] Y. J. Kim, R. G. Almond, and V. J. Shute, "Applying Evidence-Centered Design for the Development of Game-Based Assessments in Physics Playground," *International Journal of Testing*, vol. 16, no. 2, pp. 142–163, Apr. 2016, doi: 10.1080/15305058.2015.1108322.
- [12] R. Almond, R. Misley, L. Steinberg, D. Yan, and D. Williamson, "The conceptual assessment framework," in *Bayesian networks in educational assessment*, Springer, New York, 2015.
- [13] B. R. Kurniawan and M. R. A. Taqwa, "Development of Physics Problem Solving Ability Test Instrument on Dynamic Electricity Material," (in Indonesian), *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, vol. 3, no. 11, pp. 1451–1457, 2018, doi: 10.17977/jptpp.v3i11.11761.
- [14] J. Jumadi, M. I. Sukarelawan, and H. Kuswanto, "An investigation of item bias in the four-tier diagnostic test using Rasch model," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 12, no. 2, pp. 622–629, Jun. 2023, doi: 10.11591/ijere.v12i2.22845.
- [15] N. A. Sinaga, "Development of a test of problem solving and mathematical reasoning skills of junior high school students in grade VIII," (in Indonesian), *Pythagoras: Jurnal Pendidikan Matematika*, vol. 11, no. 2, Dec. 2016, doi: 10.21831/pg.v11i2.10642.
- [16] D. S. Mita, L. R. Tambunan, and N. Izzati, "Analysis of Students' Problem Solving Ability in Solving Pisa Problems," (in Indonesian), *Lentera Sriwijaya: Jurnal Ilmiah Pendidikan Matematika*, vol. 1, no. 2, p. 25, 2019, doi: 10.36706/jls.v1i2.10025.
- [17] W. Murti and T. Sunarti, "Development of a science literacy test instrument based on local wisdom in Trenggalek," (in Indonesian), *ORBITA: Jurnal Kajian, Inovasi dan Aplikasi Pendidikan Fisika*, vol. 7, no. 1, 2021, doi: 10.31764/orbita.v7i1.4386.
- [18] J. S. Barrot, I. I. Llenares, and L. S. del Rosario, "Students' online learning challenges during the pandemic and how they cope with them: The case of the Philippines," *Education and Information Technologies*, vol. 26, no. 6, pp. 7321–7338, Nov. 2021, doi: 10.1007/s10639-021-10589-x.
- [19] A. T. Capinding, "Utilization of 'Quizizz' a Game-based Assessment: An Instructional Strategy in Secondary Education Science 10," *European Journal of Educational Research*, vol. 11, no. 4, pp. 1959–1967, Oct. 2022, doi: 10.12973/eu-jer.11.4.1959.
- [20] F. A. Nieto-Escamez and M. D. Roldán-Tapia, "Gamification as Online Teaching Strategy During COVID-19: A Mini-Review," *Frontiers in Psychology*, vol. 12, May 2021, doi: 10.3389/fpsyg.2021.648552.
- [21] Z. Yildirim and M. Baran, "A comparative analysis of the effect of physical activity games and digital games on 9th grade students' achievement in physics," *Education and Information Technologies*, vol. 26, no. 1, pp. 543–563, Jan. 2021, doi: 10.1007/s10639-020-10280-7.
- [22] M. S. Darmawan, F. Daeni, and P. Listiaji, "The Use of Quizizz As An Online Assessment Application for Science Learning in The Pandemic Era," *UNNES Science Education Journal*, vol. 9, no. 3, pp. 144–150, 2020, doi: 10.15294/usej.v9i3.41541.
- [23] N. Nirwana, A. D. Sultan, and Khaeruddin, "The Application of Online Learning Assisted with Quizizz Educational Games in Learning Physics," *Jurnal Pendidikan Fisika*, vol. 9, no. 3, pp. 193–197, 2021, doi: 10.26618/jpf.v9i3.5195.

- [24] S. T. Opitz, K. Neumann, S. Bernholt, and U. Harms, "How do students understand energy in biology, chemistry, and physics? Development and validation of an assessment instrument," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 7, pp. 3019–3042, Jun. 2017, doi: 10.12973/eurasia.2017.00703a.
- [25] M. Yusup, A. Setiawan, N. Y. Rustaman, and I. Kaniawati, "Developing a Framework for the Assessment of Pre-Service Physics Teachers' Energy Literacy," *Journal of Physics: Conference Series*, vol. 877, no. 1, 2017, doi: 10.1088/1742-6596/877/1/012014.
- [26] M. Yusup, "Using Rasch model for the development and validation of energy literacy assessment instrument for prospective physics teachers," *Journal of Physics: Conference Series*, vol. 1876, no. 1, 2021, doi: 10.1088/1742-6596/1876/1/012056.
- [27] E. Etkina, D. Gitomer, C. Iaconangelo, G. Phelps, L. Seeley, and S. Vokos, "Design of an assessment to probe teachers' content knowledge for teaching: An example from energy in high school physics," *Physical Review Physics Education Research*, vol. 14, no. 1, 2018, doi: 10.1103/PhysRevPhysEducRes.14.010127.
- [28] M. Ardiyaningrum, L. Badriah, Trisniawati, Suhartini, and S. A. Widodo, "Differential item function of gender in the mathematics elementary school tryout test," *Journal of Physics: Conference Series*, vol. 1315, no. 1, p. 012036, Oct. 2019, doi: 10.1088/1742-6596/1315/1/012036.
- [29] R. Akcan and K. Atalay Kabasakal, "An Investigation of Item Bias of English Test: The Case of 2016 Year Undergraduate Placement Exam in Turkey," *International Journal of Assessment Tools in Education*, vol. 6, no. 1, pp. 48–62, Mar. 2019, doi: 10.21449/ijate.508581.
- [30] A. H. Aminudin, R. Adimayuda, I. Kaniawati, E. Suhendi, A. Samsudin, and B. Coştu, "Rasch analysis of Multitier Open-ended Light-Wave Instrument (MOLWI): Developing and assessing second-years Sundanese-scholars alternative conceptions," *Journal for the Education of Gifted Young Scientists*, vol. 7, no. 3, pp. 557–579, Sep. 2019, doi: 10.17478/jegys.574524.
- [31] B. A. Pribadi, *Design and Development of Competency Based Training Programs for Implementing the ADDIE Model*. Jakarta: Prenada Media Group (in Indonesian), 2016.
- [32] N. Novitasari, M. Ramli, and Maridi, "Development of Problem Solving Skills Assessment for High School Students on Environmental Materials," (in Indonesian), *Seminar Nasional XII Pendidikan Biologi FKIP UNS 2015*, 2015, pp. 519–525.
- [33] L. A. R. Laliyo, B. Sumintono, and C. Panigoro, "Measuring changes in hydrolysis concept of students taught by inquiry model: stacking and racking analysis techniques in Rasch model," *Heliyon*, vol. 8, no. 3, Mar. 2022, doi: 10.1016/j.heliyon.2022.e09126.
- [34] A. M. R. Tumanggor, S. Supahar, E. S. Ringo, and M. D. Harliadi, "Detecting Students' Misconception in Simple Harmonic Motion Concepts Using Four-Tier Diagnostic Test Instruments," *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, vol. 9, no. 1, pp. 21–31, Apr. 2020, doi: 10.24042/jipfalbiruni.v9i1.4571.
- [35] C. Jong, T. E. Hodges, K. D. Royal, and R. M. Welder, "Instruments to Measure Elementary Preservice Teachers' Conceptions: An Application of the Rasch Rating Scale Model," *Educational Research Quarterly*, vol. 39, no. 1, pp. 21–48, 2015.
- [36] M. J. Linacre, *A user's guide to Winsteps® Rasch-model computer program*. Beaverton, Oregon: Winsteps.com, 2016.
- [37] H. Habibi, J. Jumadi, and M. Mundilarto, "The Rasch-rating scale model to identify learning difficulties of physics students based on self-regulation skills," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 8, no. 4, pp. 659–665, Dec. 2019, doi: 10.11591/ijere.v8i4.20292.
- [38] Firstananda, Erba, and Sugiono, "Development of Mathematics Learning Tools for Circle Materials with Guided Discovery Approach for Class VIII Junior High School Students," (in Indonesian), *Journal of Chemical Information and Modeling*, vol. 53, no. 9, pp. 1689–1699, 2019.
- [39] B. Sumintono and W. Widhiarso, *Application of Rasch Model for Social Sciences Research*. Trim Komunikata Publishing House (in Indonesian), 2014.
- [40] S. M. Ayu, E. Gustina, L. Sofiana, Y. Wardani, and M. I. Sukarelawan, "Physical and psychological violence victimization scale in adolescents dating: Confirmatory factor analysis and Rasch model," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 12, no. 1, pp. 96–105, Mar. 2023, doi: 10.11591/ijere.v12i1.22250.
- [41] M. Ibnu, B. Indriyani, H. Inayatullah, and Y. Guntara, "Application of Rasch model: Development of a test instrument to measure student misconceptions," (in Indonesian), *Prosiding Seminar Nasional Pendidikan FKIP*, vol. 2, no. 1, 2019, pp. 205–210.
- [42] W. J. Boone, M. S. Yale, and J. R. Staver, *Rasch analysis in the human sciences*. Dordrecht: Springer Netherlands, 2014. doi: 10.1007/978-94-007-6857-4.
- [43] B. Setiawan, M. Panduwangi, and B. Sumintono, "A Rasch analysis of the community's preference for different attributes of Islamic banks in Indonesia," *International Journal of Social Economics*, vol. 45, no. 12, pp. 1647–1662, 2018, doi: 10.1108/IJSE-07-2017-0294.
- [44] M. R. Espejo, "Applying the Rasch Model: Fundamental Measurement in the Human Sciences," *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 167, no. 2, pp. 377–377, 2004, doi: 10.1111/j.1467-985x.2004.t01-2-.x.
- [45] N. Papini, M. Kang, S. Ryu, E. Griese, T. Wingert, and S. Herrmann, "Rasch calibration of the 25-item Connor-Davidson Resilience Scale," *Journal of Health Psychology*, vol. 26, no. 11, pp. 1976–1987, Sep. 2021, doi: 10.1177/1359105320904769.
- [46] T. G. Bond and C. M. Fox, *Applying the Rasch model: Fundamental measurement in human sciences*, New Jersey: Lawrence Erlbaum Associates, 2007.
- [47] K. Fenditasari, Jumadi, E. Istiyono, and Hendra, "Identification of misconceptions on heat and temperature among physics education students using four-tier diagnostic test," *Journal of Physics: Conference Series*, vol. 1470, no. 1, p. 012055, Feb. 2020, doi: 10.1088/1742-6596/1470/1/012055.
- [48] M. Maison, I. C. Safitri, and R. W. Wardana, "Identification of Misconception of High School Students on Temperature and Calor Topic Using Four-Tier Diagnostic Instrument," *Edusains*, vol. 11, no. 2, pp. 195–202, Jan. 2020, doi: 10.15408/es.v11i2.11465.
- [49] J. I. Utari and F. U. Ermawati, "Development of a Four-Tier Misconception Diagnostic Test Instrument for Temperature, Heat and Its Transfer Materials," (in Indonesian), *Inovasi Pendidikan Fisika*, vol. 7, no. 3, pp. 434–439, 2018.
- [50] D. O. Göksün and G. Gürsoy, "Comparing success and engagement in gamified learning experiences via Kahoot and Quizizz," *Computers and Education*, vol. 135, pp. 15–29, Jul. 2019, doi: 10.1016/j.compedu.2019.02.015.
- [51] S. Zuhriyah and B. W. Pratolo, "Exploring students' views in the use of Quizizz as an assessment tool in English as a foreign language (EFL) class," *Universal Journal of Educational Research*, vol. 8, no. 11, 2020, doi: 10.13189/ujer.2020.081132.
- [52] M. L. H. Abbas, "Development of Computer Based Diagnostic Test for Student Misconception on Material Temperature and Heat," *Jurnal Pendidikan Fisika dan Keilmuan (JPFK)*, vol. 6, no. 1, p. 12, Mar. 2020, doi: 10.25273/jpfk.v6i1.5153.
- [53] I. Kaniawati, A. Danawan, I. Suyana, A. Samsudin, and E. Suhendi, "Implementation of Interactive Conceptual Instruction (ICI) With Computer Simulation: Impact of Students' Misconceptions on Momentum and Impulse Material," *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, vol. 10, no. 1, pp. 1–17, Apr. 2021, doi: 10.24042/jipfalbiruni.v10i1.8375.
- [54] C. C. Chou, "An analysis of the 3D video and interactive response approach effects on the science remedial teaching for fourth grade underachieving students," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 4, pp. 1059–1073, 2017, doi: 10.12973/eurasia.2017.00658a.




- [55] Y. Kong, L. R. Kayumova, and V. G. Zakirova, "Simulation technologies in preparing teachers to deal with risks," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 8, p. 4753, 2017, doi: 10.12973/eurasia.2017.00962a.
- [56] E. Sanyoto, "Application of Interactive Demonstration Learning Models Assisted by Virtual Simulation Media to Reduce Students' Misconceptions on Temperature, Heat, and Heat Transfer Materials," (in Indonesian), *Inovasi Pendidikan Fisika*, vol. 5, no. 3, pp. 379–389, 2017.
- [57] J. S. Lai and D. T. Eton, "Clinically meaningful gaps," *Rasch Measurement Transactions*, vol. 15, no. 4, p. 850, 2002.

BIOGRAPHIES OF AUTHORS






Khoirun Nisa'    is a student at the Department of Physics, Faculty of Mathematics and Natural Science, Universitas Negeri Surabaya (UNESA). Her research interests in assessment, physics education, and cultural studies. She can be contacted at email: khoirun19005@gmail.com.






Nadi Suprpto    is a Professor at the Department of Physics, Faculty of Mathematics and Natural Science, Universitas Negeri Surabaya (UNESA). His research interests in the physics literacy, physics education, 21st Century teaching and learning, assessment, cultural studies, curriculum, and philosophy of science. He can be contacted at email: nadisuprpto@unesa.ac.id.






Mohd Zaidi Bin Amiruddin    is a student in the Department of Physics, Faculty of Mathematics and Natural Science, Universitas Negeri Surabaya, Indonesia. He has a solid intention to continue his education at a higher level. He can be contacted at email: mohdzaidi.19079@mhs.unesa.ac.id.



Eka Putri Dian Nata Sari    is the fourth-year undergraduate student in the Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, Surabaya, Indonesia. She can be contacted at email: eka.19053@mhs.unesa.ac.id.



Bilqisth Dwi' Athiah    is the second-year undergraduate student in the Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, Surabaya, Indonesia. She can be contacted at email: bilqisthdwi.21068@mhs.unesa.ac.id.