



This work is licensed under

a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

Analysis of the Quality of the Formative Test Items for Physics Learning Using the Rasch Model in the 21st Century Learning

Muhammad Asriadi¹, Samsul Hadi²
Yogyakarta State University, Sleman, Indonesia^{1,2}
asriadi190197@gmail.com¹, samsul_hd@uny.ac.id²

Received: October 27th, 2020. Revised: December 19th, 2020. Accepted: January 12th, 2021

Keywords :

21st Century Learning;
Formative Test; Physics
Learning; Rasch Model

ABSTRACT

Tests that have good quality are tests that have an even level of difficulty and can be completed as a whole by every respondent with every level of ability. This article aims to identify and analyze the quality of the items from the formative test material for static and dynamic fluids in physics subjects. This research used a survey with a cross-sectional method. Convenience sampling is a sampling technique used in this research. Then the sample in this study were 52 high school students class XI. The research instrument used a formative test for physics subjects, which were then analyzed using the Rasch model with the help of Winsteps software. The results of this study show that this formative test has a reliability level of 0.88 (very good) with a good level of problem difficulty and a good item fit level compatibility and there are no biased items in measurement so that this formative test is feasible to be used as a standard item in measuring students' abilities in materials of static and dynamic fluid in physics subjects.

INTRODUCTION

In recent decades, 21st century skills assessment has become a topic of concern in education [1]. In 21st century learning emphasizes increasing student competence. Learners in the 21st century should be facilitated to experience the best learning experiences to achieve learning goals effectively [2]. To measure students' ability in competency engagement, an assessment that has good quality is needed [3]. Assessment in the form of a test is part of the activity to measure the abilities of students. Thus the assessment will be useful to see the overall quality of education and this assessment will also provide important information for improving the learning process [4]. In the 2013 curriculum, the assessment of learning outcomes with authentic assessment uses to test and non-test techniques, performance, and attitude assessments [5]. The assessment of learning outcomes is used by educators on an ongoing basis to assess the achievement of student competencies, materials for compiling reports on learning outcomes, and improving the learning process, especially in learning physics.

Physics is one part of learning science. In science learning includes observing, measuring, classifying, concluding, predicting, hypothesizing, identifying variables, experimenting, and interpreting data [6]. It all boils down to problem-solving abilities. Learning physics must be able to facilitate students to be

able to develop problem-solving skills, this affects the motivation of students to learn and develop the skills needed [7]. In addition, learning physics requires active students so that it will strengthen their understanding of physics concepts [8]. This is following the principles of constructivism, where knowledge is built by students themselves where students who are not active, cannot receive knowledge directly transferred by the teacher. Strategies that can be used to assist students in achieving completeness of physics material are by giving a test that can measure the level of understanding and ability of students [9].

Formative tests are carried out by teachers who provide learning material, interact directly, and see student progress over time. Formative evaluation to obtain information about students is carried out by giving assignments, written tests, questions, and answers in class, and various other forms of evaluation [10]. Formative assessment allows students to become aware of their difficulties, and perhaps, to recognize and correct their own mistakes [11]. For evaluation effectiveness, the form of assessment needs to be adjusted to student characteristics, such as intelligence, cognitive styles, learning styles, work attitudes, talents, and others [12]. In this connection, the aspects of using the test are designed following the abilities of students who answer them [13]. So that in this formative test, the assessment applies the item response theory with the Rasch model in determining the quality of the assessment of participant physics learning outcomes.

Item response theory considers the test taker's behavior at the item level. The assessment of the IRT model involves collecting various kinds of diagnostic evidence for the suitability of the model and then making informed judgments about the fit of the model and the usefulness of the model with a particular data set [14]. Item response theory rests on two basic postulates, namely: (1) the performance of the examinee on each test item can be predicted by a set of factors called latent or trait traits; and (2) the relationship between test item performance and the set of traits underlying the item performance can be explained by a monotonically increasing function called the item character function or item characteristic curve (ICC) [15]. The item response model determines the relationship between the examinee's observable abilities and unobservable traits or assumed abilities underlying test performance [16].

One of the modelings of grain response theory is the Rasch model. The Rasch model performs a thorough analysis of students' ability to solve problems. Rasch reads the pattern of an event so that it is predictive of its ability to solve missing data problems [17]. The main assumption of the Rasch model is that the assessment of the response item categories must be the same where the score must continue to increase according to the level of difficulty [18]. The main characteristic of the Rasch model is that it considers all responses from a test taker regardless of the order in which the problem was solved [19]. That means the level of difficulty of each test item does not have to be in consecutive order. The main advantage of the Rasch model is the process used by participants in solving problems more accurately. Each person's score is determined by the interaction between person size, item difficulty, and the score category threshold [20]. Rasch modeling can produce standard error measurement values which can improve the accuracy of calculations [21]. In the Rasch model, the probability of students being randomly selected at the ability level (θ) to get the correct answer for each item (i) can be stated as the following formula.

$$P_i(\theta) = \frac{1}{1+e^{-D(\theta-b_i)}} \quad (1)$$

Information:

e : exponential constant = 2.718

D : scale factor = 1.7

The results of the Rasch model show that the level of students' ability to do the questions is very dependent on the level of ability and difficulty level of the questions [22]. So that the probability of success is the ability of the respondent minus the difficulty level of the item. One of the advantages of applying the Rasch model to analyze assessment data is its ability to consider and analyze assessment

items and students as test takers and then present the results together. [23]. So that the Rasch model can provide an objective measurement based on student abilities that do not depend on the difficulty of the item in the assessment task. Based on the description above, the purpose of writing this article is to identify and analyze the quality of the items from the formative test material for static and dynamic fluids in physics subjects.

METHOD

This study was a survey with a *cross-sectional survey* design. The *cross-sectional survey* collects information from a sample drawn from a predetermined population. Subsequently, the information is collected at one point in time [24]. A *cross-sectional study* can examine current attitudes, beliefs, opinions, or practices. Attitudes, beliefs, and opinions are the entry points for what individuals think about problems, whereas practice is their actual behavior [25]. The population is students of class XI in high schools in Indonesia taking physics subjects in the academic year 2020/2021 who are willing to become respondents. The sampling technique used is *convenience sampling*. In *convenience sampling*, the researcher selects participants because they are willing and available for study [25]. Then there were 52 high school class XI students consisting of 23 man and 29 woman who were willing to be samples to do the formative test questions. The instrument used was a formative test on physics subjects with static and dynamic fluids. The answers to the test questions that have been done by students are then analyzed in terms of the quality of the questions in terms of difficulty level, item fit level and differential item functioning using the Rasch model based on the theory of question response with the help of WINSTEPS version 3.73.

RESULTS AND DISCUSSIONS

Formative Test Construction Map

The type of test in this study is a formative test with static fluid and dynamic fluid. This material is studied in class XI semester 2. The number of questions is 20 items, with a description of the comparisons as shown in the table below.

Table 1. Problem Grids

Subject : Physics
 Class / Semester : XI / II
 Basic Competence : Analyzing laws related to static and dynamic fluids and their application in everyday life

Syllabus Indicators	Problem Indicators	Cognitive Realm	Number of Questions
Formulate the basic law of static fluid	1. Students can explain the surface height of a liquid-based on the main laws of thermodynamics correctly	C2	1
	2. Students can determine the height of the substance in the U pipe correctly	C3	2
	3. Students can correctly analyze the factors that influence the amount of hydrostatic pressure	C4	3
	4. Students can determine the amount of force that must be applied to lift a load according to Pascall's law correctly	C3	2

Syllabus Indicators	Problem Indicators	Cognitive Realm	Number of Questions
Apply the basic laws of static fluids to everyday physics problems	5. Students can calculate the volume of objects in the event of float correctly	C3	2
	6. Students can determine the amount of upward force in the floating event correctly	C3	2
	7. Students can determine the volume of objects in the floating event correctly	C3	2
	8. Students can determine the density of objects in the hover event correctly	C3	2
Formulate the basic law of dynamic fluids	9. Presented data volume, cross-sectional area, speed, students can determine the amount of time needed to fill the tub with water correctly	C3	2
	10. A picture of the fluid flowing in a pipe with a different cross-sectional area is presented. Students can correctly analyze the flow rate and flow rate	C4	2

The grid above is arranged based on the syllabus that has been established by the government in the 2013 curriculum and each item indicator is adjusted to a cognitive level according to the bloom taxonomy. In 21st century learning, it is more important to master competence in terms of tying the ability to think critically, creatively and find solutions to problems [26]. The test question grid is designed in such a way as to be able to increase the level of their thinking skills and competencies, especially in static and dynamic fluid material.

Summary Statistic

Winsteps provides general item analysis in the information on summary statistics of the test presented. To find out the value can be seen in the table below.

Table 2. Summary Statistics

Statistical Component	Value
Item Questions	20
Test Participants	52
Log-Likelihood Chi-Square	949.09 with 950 d.f. p = 0.5023
Average Measurement	-0.66
Cronbach alpha (KR-20) Person Raw	0.88
Score "Test" reliability	0.82
Person Reliability	0.86
Item Reliability	2,48

Based on the table 2 above, information is obtained that the data processed were 52 test participants with 20 item questions. The resulting Chi-Square value is 949.09 with 950 degrees of freedom and $p = 0.5023$ where $p > 0.01$. This shows that in general the measurements taken are quite good and the results are significant. Besides, the average measurement value obtained at the output person is -0.66 ($\mu < 0.00$). A negative score indicates that the student is having problems answering the question. An average score that is smaller than 0 indicates that the tendency of the test taker's ability is smaller than the difficulty level of the questions.

The reliability value with Cronbach alpha (KR-20) which measures the interaction between the person and the item as a whole is $\alpha = 0.88$, this shows that there is a match between the respondent and the instrument used. It also shows that the level of precision and consistency of test scores is good. With

known reliability values ranging from 0-1. The higher the reliability coefficient of a test (close to 1), the higher the accuracy [27]. This is evidenced by the Alpha coefficient value of 0.88. Thus this test has a high-reliability category. A reliable measuring tool consists of valid items. Besides, it is known that the person's reliability value is 0.82 and the item reliability value is 0.86. This shows that the consistency of the answers from the test takers is strong and the quality of the items on the reliability aspect of the test is categorized as good. There is also a separation value that indicates the quality of the test and the quality of the test taker. The greater the separation value the better because it can identify a wider group of subjects (capable or unable) and a group of questions (difficult or easy) [28]. The formula that can be used to see the grouping more accurately is called strata separation with the formula $H = [(4 \times \text{separation}) + 1] / 3$. The result of the winsteps output is as shown in Figure 2, it is known that the value of item separation is 2.48 then $H = [(4 \times 2.48) + 1] / 3$, namely 3.64 or rounded to 4. This indicates that the respondent only consists of four groups. More splits indicate that the test is good. The higher the grain separation value, the better the measurement. Based on these results it is known that all test items have good quality as evidenced by the reliability value which is in the good category. Of course this is very supportive in terms of measuring students' abilities. This is because one of the main keys in assessing the success of learning in the 21st century starts with quality test questions [29]. So that's way, this formative test question can be used as a measuring tool in assessing students' abilities.

Person Map Item

Rasch's analysis provides a person-item distribution map known as the Wright Map which is nothing but a person-item map. The Wright map describes the distribution of test-takers' abilities and the distribution of the difficulty level of items with the same scale and provides an overview of the readiness of the respondents by placing the difficulty level of the test on the same measurement scale as the respondent's ability [30]. The output of the person-item map can be seen in the image below.

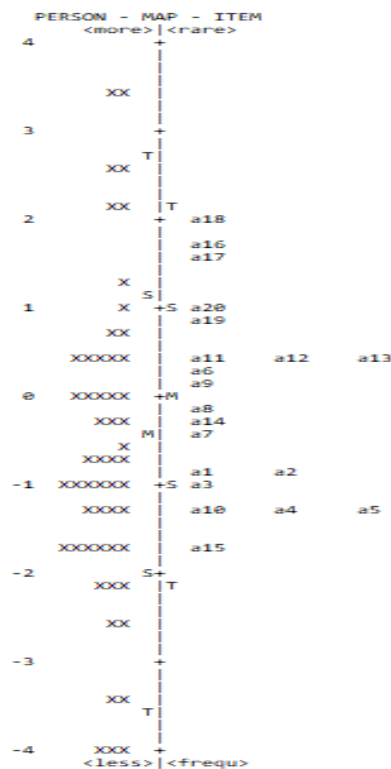


Fig 1. Person Map Item

In figure 1 above, it can be seen that on the left side is the distribution of subject abilities, while on the right is the distribution of items. From this map, it can be seen that in general, the questions on the test are equivalent when compared to the ability of the subjects. This means that the 20 test items can be

worked on by the test takers. There are no items that are the most difficult. All questions are in the medium difficulty category. Besides, all questions are in the interval -2 to 2, which means that the questions given to students do not have a moderate or good level of difficulty. So that the formative test questions are good enough to use. From these findings it can be concluded that the level of student ability is only in the medium category. This shows that learning physics in the classroom still needs to be improved in facing the rate of technological development, social and economic changes at the global level [31]. So that students' abilities are not enough to be in the medium category.

Item Difficulty Level

One thing we need to pay attention to is the results of the Rasch analysis with these winsteps. A high logit (measure) value indicates that the item has a high difficulty level. This correlates with the total score, where the few correct answers in the total score correlate with higher measure scores. This data measure also has the same scale. To find out the classification of the level of difficulty of each item can be seen in table 3 below.

Table 3. Level of Difficulty Item

Item	Level of Difficulty	Description	Item	Level of Difficulty	Description
1	-0.80	Easy	11	0.47	Difficult
2	-0.91	Easy	12	0.47	Difficult
3	-1.02	Very Easy	13	0.47	Difficult
4	-1.25	Very Easy	14	-0,25	Easy
5	-1.25	Very Easy	15	-1.72	Very Easy
6	0.34	Difficult	16	1.73	Very Difficult
7	-0.36	Easy	17	1.53	Very Difficult
8	-0.14	Easy	18	1.93	Very Difficult
9	0.10	Difficult	19	0.88	Difficult
10	-1.25	Very Easy	20	1.03	Very Difficult

Based on table 3, it will be classified into a category table for the difficulty level of the questions to understand the percentage of questions. The order of the hardest questions in question 18 and the easiest question 15. To see the percentage of questions can be seen in Table 4 below.

Table 4. Category Level of Difficulty Item

Category	Frequency	Percentage (%)
Very Easy	5	25
Easy	5	25
Difficult	6	30
Very Difficult	4	20
Total	20	100

Based on table 4 above, it is known that most of the items used as formative tests have a proportional difficulty level. So that the questions made have met the ideal criteria to be used as a formative test of static and dynamic fluid material.

Item Fit Level

The suitability level of this item is used to see the accuracy of the item with the model or fit item. Item fit level explains whether our item has a normal function to take measurements or not. If there are items that are not fit, this indicates a misconception of the subject in answering the question [32]. To find out the difficulty level of an item, it can be seen in the table below.

Table 5. Item Statistics: Misfit Order

ITEM	INFIT		OUTFIT		PT-MEASURE	
	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.
18	2.19	3.1	3.01	2.1	.01	.49
16	1.43	1.4	1.89	1.3	.31	.50
9	1.37	2.1	1.46	1.4	.39	.55
7	1.11	.8	1.34	1.1	.49	.56
3	1.16	1.2	1.08	.3	.49	.55
17	.80	-.7	1.04	.3	.56	.51
14	1.01	.2	.95	-.1	.55	.56
5	1.00	.0	.89	-.1	.55	.55
11	.98	-.1	.85	-.4	.57	.55
10	.97	-.2	.78	-.4	.57	.55
4	.97	-.2	.86	-.2	.56	.55
2	.94	-.4	.80	-.5	.59	.55
20	.93	-.2	.86	-.2	.56	.53
15	.93	-.4	.72	-.4	.58	.54
12	.90	-.5	.76	-.6	.60	.55
13	.87	-.7	.71	-.8	.61	.55
19	.74	-1.3	.55	-1.2	.65	.54
1	.74	-2.1	.62	-1.2	.67	.55
6	.68	-2.0	.59	-1.4	.68	.55
8	.67	-2.4	.53	-1.8	.70	.56
MEAN	1.02	-.1	1.02	-.1		
S.D	.33	1.3	.56	1.0		

Based on table 5 above, it provides information about the value of outfit means-square, outfit z-standard, and point measure correlation which are the criteria used to see the level of suitability of items. If the item does not meet the criteria, it should be replaced or repaired. The guide for assessing item suitability criteria is the accepted Outfit Mean Square (MNSQ) value: $0.5 < \text{MNSQ} < 1.5$. Accepted standard Z clothing value (ZSTD): $-2.0 < \text{ZSTD} < +2.0$ [33]. Received Point Size Correlation Value: $0.4 < \text{pt measure correlation} < 0.85$. In addition, the output has a value of Point Measure Correlation which is classified as very good (> 0.40), good ($0.30 - 0.39$), sufficient ($0.20 - 0.29$), unable to differentiate ($0.00 - 0.19$), and requires an item inspection (< 0.00) [34]. Winsteps has sorted the items based on which items do not match. Items that don't fit properly are placed at the top. Also, on the Mean Square (MNSQ) output and Z standard outfit (ZSTD), all items have met the fit criteria. The item displayed has a Point Measure Correlation value, there is only 1 item that cannot be biased, namely item number 18. Meanwhile, the other nineteen items are classified as very good. So it is suggested that the 18 item must be reviewed and corrected in terms of substance.

Differential Item Functioning

In the winsteps program package, information about this item bias can be seen through Item: DIF, between / within. Items that have a P value ($\text{PROB} < 0.05$) indicates that the item is infected with DIF. In the results of the winsteps analysis, it is known that the overall probability value of the items is 1,000 ($\text{PROB} > 0.05$), which means that all items in the formative test of static and dynamic fluid do not experience bias in their measurements.

CONCLUSION AND SUGGESTION

In this study, it shows that the formative test has met the valid and reliable requirements based on the Rasch model. From the results of identification and analysis using winsteps this formative test has a good level of problem difficulty, a good level of fit item level, and there are no items that are biased in measurement. As a suggestion for the need for the participation of teachers in producing test questions that have a good level of quality in measuring student competencies, especially in 21st century learning

which emphasizes the binding of 21st century life skills competencies. With good tests, the teacher will easily identify the abilities of each student. Also, the benefits that can be obtained from this study are obtained by analyzing the quality of a test instrument with the Rasch model which can be a guide for teachers to produce test instruments that have good quality in measuring student competence in 21st century learning.

ACKNOWLEDGEMENTS

The author would like to thank all class XI high school students who have been willing to be respondents in this case doing the formative tests that the author has made. Hopefully, from the results of this research, formative tests can be useful, especially in learning physics.

REFERENCES

- [1] Haviz, M., Maris, I. M., & Fudholi, A. (2020). Assessing Pre-Service Teachers' Perception on 21st century Skills in Indonesia. *Journal of Turkish Science Education*, 17(3): 351-363.
- [2] Herbert, Putro, B. L., Putra, R. R. J., & Fitriyani, N. S. (2019, November). Learning Management System (LMS) model based on machine learning supports 21st century learning as the implementation of curriculum 2013. In *Journal of Physics: Conference Series* (Vol. 1280, No. 3, p. 032032). IOP Publishing.
- [3] Arding, N. I., & Atun, S. (2020, January). Analysis of Junior High School students' scientific literacy on simple effort and aircraft for everyday life. In *Journal of Physics: Conference Series* (Vol. 1440, No. 1, p. 012095). IOP Publishing.
- [4] Ofianto, O., & Suhartono, S. (2015). An assessment model of historical thinking skills by means of the RASCH model. *REiD (Research and Evaluation in Education)*, 1(1): 73-83.
- [5] Anikarnisia, N. M., & Wilujeng, I. (2020, January). Need assessment of STEM education based on local wisdom in junior high school. In *Journal of Physics: Conference Series* (Vol. 1440, No. 1, p. 012092). IOP Publishing.
- [6] Subali, B., Kumaidi, K., Aminah, N. S., & Sumintono, B. (2019). Student achievement based on the use of scientific method in the natural science subject in elementary school. *Jurnal Pendidikan IPA Indonesia*, 8(1): 39-51.
- [7] Gebze, D. A., & Perwati, S. (2020, January). Improving problem-solving ability in physics through android-based mobile learning application. In *Journal of Physics: Conference Series* (Vol. 1440, No. 1, p. 012022). IOP Publishing.
- [8] Istiyono, E., Dwandaru, W. B., & Rahayu, F. (2018). The developing of creative thinking skills test based on modern test theory in physics of senior high schools. *Cakrawala Pendidikan*, 3(2).
- [9] Larasati, P. E., & Yunanta, D. R. A. (2020, January). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. In *Journal of Physics: Conference Series* (Vol. 1440, No. 1, p. 012020). IOP Publishing.
- [10] Nurjannah, N. (2017). Efektivitas Bentuk Penilaian Formatif Disesuaikan Dengan Media Pembelajaran. *PARAMETER: Jurnal Pendidikan Universitas Negeri Jakarta*, 29(1): 75-90.
- [11] da Silva, P. N., & de Abreu, R. C. P. (2017). DIA: A Computerized Adaptive Testing Tool for Assessing Student Learning.
- [12] Jensen, N., Rice, A., & Soland, J. (2018). The influence of rapidly guessed item responses on teacher value-added estimates: Implications for policy and practice. *Educational Evaluation and Policy Analysis*, 40(2): 267-284.
- [13] De Silva, C. R., & Hill, M. A. (2013). Higher order reading skills and reader response theory: Strategies for the classroom.
- [14] Osarumwense, J. H., & Duru, C. P. (2019). Assessment Of Model Fit For 2016 And 2017 Biology Multiple Choice Test Items Of The National Business And Technical Examination Board. *International Journal of Innovation Education and Research*, 7(4): 12-22.
- [15] Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response*

- theory. Sage.
- [16] Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.
- [17] Abdullah, H., Arsad, N., Hashim, F. H., Aziz, N. A., Amin, N., & Ali, S. H. (2012). Evaluation of students' achievement in the final exam questions for microelectronic (KKKL3054) using the Rasch model. *Procedia-Social and Behavioral Sciences*, 60: 119-123.
- [18] Habibi, H., Jumadi, J., & Mundilarto, M. (2019). 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*, 8(4): 659-665.
- [19] Isnani, I., Utami, W. B., Susongko, P., & Lestiani, H. T. (2019). Estimation of college students' ability on real analysis course using Rasch model. *REiD (Research and Evaluation in Education)*, 5(2): 95-102.
- [20] Mahmud, Z., & Porter, A. (2015). Using Rasch Analysis to Explore What Students Learn about Probability Concepts. *Indonesian Mathematical Society Journal on Mathematics Education*, 6(1): 1-10.
- [21] Lia, R. M., Rusilowati, A., & Isnaeni, W. (2020). NGSS-oriented chemistry test instruments: Validity and reliability analysis with the Rasch model. *REiD (Research and Evaluation in Education)*, 6(1): 41-50.
- [22] van der Linden, W. J., & Hambleton, R. K. (Eds.). (2013). *Handbook of modern item response theory*. Springer Science & Business Media.
- [23] Bergh, D. (2020). Rasch Measurement: Applications in Quantitative Educational Research.
- [24] Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (1993). *How to design and evaluate research in education* (Vol. 7). New York: McGraw-Hill.
- [25] Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative* (pp. 146-166). Upper Saddle River, NJ: Prentice Hall.
- [26] Al Kandari, A. M., & Al Qattan, M. M. (2020). E-Task-Based Learning Approach to Enhancing 21st-Century Learning Outcomes. *International Journal of Instruction*, 13(1): 551-566.
- [27] Purnamasari, U. D., & Kartowagiran, B. (2019). Application rasch model using R program in analyze the characteristics of chemical items. *Jurnal Inovasi Pendidikan IPA*, 5(2): 147-156.
- [28] Wibisono, S. (2019). Aplikasi model rasch untuk validasi instrumen pengukuran fundamentalisme agama bagi responden muslim. *JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)*, 3(3).
- [29] Ujir, H., Salleh, S. F., Marzuki, A. S. W., Hashim, H. F., & Alias, A. A. (2020). Teaching Workload in 21st Century Higher Education Learning Setting. *International Journal of Evaluation and Research in Education*, 9(1): 221-227.
- [30] Salman, A., & Aziz, A. A. (2015). Evaluating user readiness towards digital society: a Rasch measurement model analysis. *Procedia Computer Science*, 65: 1154-1159.
- [31] Göçen, A., Eral, S. H., & Bücü, M. H. (2020). Teacher Perceptions of a 21st Century Classroom. *International Journal of Contemporary Educational Research*, 7(1): 85-98.
- [32] Camminatiello, I., Gallo, M., & Menini, T. (2010). The Rasch Model for Evaluating Italian Student Performance. *Journal of Applied Quantitative Methods*, 5(2): 331-349.
- [33] Susongko, P. (2016). Validation of science achievement test with the rasch model. *Jurnal Pendidikan IPA Indonesia*, 5(2): 268-277.
- [34] Boone, W. J., Staver, J. R., & Yale, M. S. (2013). *Rasch analysis in the human sciences*. Springer Science & Business Media.