JPF | Volume 11 | Number 1 | 2023 | 59 - 72

p - ISSN: 2302-8939
e - ISSN: 2527-4015

Jurnal Pendidikan Fisika

https://journal.unismuh.ac.id/index.php/jpf

DOI: 10.26618/jpf.v11i1.9228

Development of Questionnaire to Evaluate Students' Perception about Real and Virtual Refutational Laboratory: A Rasch Measurement Approach

Andi Moh Amin[1)*], Andi Suhandi[2)], Ridwan Efendi[3)]

[1),2),3)]Physiscs Education Department, Universitas Pendidikan Indonesia, Bandung, 401564, Indonesia

**Corresponding author: andimoh_amin@upi.edu*

Received: October 15, 2022; Accepted: January 02, 2023; Published: January 31, 2023

Abstract – *This study aims to develop the questionnaire of students' perception about the real and virtual refutational laboratory which is a misconception remediation method adapted from refutation text using a new approach i.e., laboratory. This study was a quantitative description that involved 53 students (12 males and 41 females) chosen using purposive sampling. The students were initially given a trial questionnaire after receiving a misconception remediation treatment using a real and virtual laboratory method in different classes. The validity and reliability of the instrument were analyzed using Rasch Model. The analysis result reveals that the Cronbach's alpha was 0.61 (enough category), the item reliability was 0.82 (good category) and the person reliability was 0.31 (less variance in students' answers). Meanwhile, based on the validity test, the raw variance was obtained from measures of 29.4% reaching the minimum criteria range of 20%. The Differential Item Function (DIF) test based on gender results revealed that an item was biased (p=0.0435), whereas, there is no biased instrument based on the kind of class in the Differential Item Function test study. The analysis demonstrates that after correcting and deleting some items that did not match based on its criteria, this instrument may be used to gather information on students' perception regarding the implementation of laboratory learning methods or the like in a valid and reliable way by the researchers and teachers.*

Keywords: *misconception; questionnaire; rasch model; real laboratory; virtual laboratory*

I. INTRODUCTION

The covid-19 has had a major impact on education (Saputra et al., 2021) due to the learning process that cannot run optimally, such as online learning and learning shortly so that in learning physics concepts or other subjects a teacher cannot teach in depth which will have an impact on the learning process in understanding of students' concepts (Cui & Yu, 2019). It was because online learning using media can not support interactive learning (Setiyoaji et al., 2021). In addition, online learning will cause students to focus less on learning, feel bored during learning (Rasmitadila et al., 2020; Yunitasari & Hanifah, 2020) and lack enthusiasm in participating in learning (Sintema, 2020) and also have an impact on student's psychology. A study conducted by Sinatra, (2022) stated that motivation and emotions will have an impact on misconceptions. This is because low motivation causes a person not to be interested in digging for in-depth information and not be skeptical of the information he receives (Marton & Pang, 2008).

Another thing is that not all students have gadgets or laptops and poor internet connections, so students only study with the help of books which has an impact on the emergence of misconceptions about the teaching materials being studied because they do not get optimal guidance (Liu & Fang, 2016). Marton & Pang (2008) writes that there are several student failures in understanding science and mathematics concepts, namely low critical thinking skills, knowledge fragmentation, lack of knowledge transfer and misconceptions. This problem mainly arises when it is associated with cases where the new information to be obtained contradicts information already possessed by previous students or information that was previously believed in the environment but the concept is not scientific.

Misconceptions are resistant to new scientific concepts so students who experience misconceptions will find it difficult to accept new information even though the concept is more scientific (Posner et al., 1982). Yet, to reduce their conception, cognitive conflict is needed because it is effective way for helping students' conception (Haryono et al., 2021). So that one approach that can be taken in remediating students' misconceptions is by using the refulational laboratory method; this method was adapted from the refutational tax method of learning but with a laboratory approach as an accommodation in remediating students' misconceptions.

Misconceptions can be an obstacle to the entry of new concepts so misconceptions must be identified and cured immediately (Tumanggor et al., 2020). One part of the conceptual change model is the refutation text which is not only developed in reading literacy education but also in teaching science including physics.

Refutation text can be done by activating these misconceptions by creating conflict in order that there will be doubts in students about the concepts they have. In these conditions, students can be given alternative concepts that can be understood and make sense as well as more rational evidence (Lem et al., 2017).

In increasing students' understanding of concepts, it is important to involve the role of laboratories in learning. Laboratory activities have a role to connect concepts and the material studied by students. Laboratory activities can also increase students' interest in learning. In addition, learning in this way can provide direct experience (Hodson, 1996) in science learning so to crease students' understanding and interest (Harman et al., 2016).

Based on these advantages, a remediation learning method was adapted using a laboratory approach called the refutational laboratory. There are two modes of learning can direct learning in a laboratory named "Real refutational Laboratory" and learning carried out using visual media or simulators called "Virtual refutational Laboratory".

The refutational method using a laboratory approach is a remediation method of misconceptions that is still underdeveloped. Therefore in conducting research or learning, researchers or teachers need instruments to evaluate the implementation of remediation of misconceptions that use the refutational laboratory method as a form of educational improvement quality (Cuadros et al., 2021; Lutasari & Kartowagran, 2019).

However, in fact the researchers did not find any supporting instruments, so in this study the researcher designed a non-test instrument that can be used for collecting, analyzing, and presenting information about students' learning with this method in the form of a questionnaire. It was developed to assess the effectiveness of using a learning method in remediating misconceptions. This misconception instrument was developed by adapting several related instruments so that it is used as a valid and reliable questionnaire instrument.

As indicator that an instrument is feasible to use is to look at the validity and reliability which is carried out using the Rasch model and assisted by Winistap software version 5.2.5.1 (Sumintono & Widhiarso, 2014). This model has two remarkable properties of invariant and interval scaling which is achieved from the basic assumption of usni-dimensional i.e. when the data fit the model (Tabatabaee-Yazdi et al., 2018). Rasch measurement is based on the interaction between items and persons described based on mathematical equations. People who have high abilities must answer questions with an easier level of difficulty (Soeharto, 2021; Sumintono & Widhiarso, 2014).

Measurement using the Rasch item in a certain sense is an ideal measurement because the measurement is valid, objective and adequate. An analysis using the Rasch model has two different objectives. The first is to provide a calibration equation that relates the total score on the Education test with the estimated value of the underlying latent trait variable of the respondent's response that is characterized as a person measurement procedure. The second objective of item analysis using the Rasch model is scale validation. Measurement using the Rasch model is the simplest item response model in terms of mathematics and statistics (Kreiner, 2007).

In addition to the refutational laboratory questionnaire design, the use of the Rusch model in analyzing the validity and reliability of the item questions is one of the novelties in this study because there are not many similar articles that use Rasch analysis in assessing the feasibility of their instruments.

This paper aims to evaluate and design a learning questionnaire that is valid and reliable with using the refutational lab method to obtain an overview that can be used in exploring the experiences felt by respondents in remediating misconceptions. In this article, some related questions are raised to answer the validity and the reliability in this instrument. (1) Does the questionnaire developed reliable based on Rasch measurement? (2) Does the questionnaire developed valid based on Rasch measurement? (3) Are there items that are biased towards gender and class type? and (4) How do items and persons interact in the development questionnaire?

II. METHODS

This study was designed to produce real and virtual refutational laboratory instrument products by adopting the Borg and Gall models. There are three important stages carried out to produce a good instrument product, while the three steps are: (1) the research and collecting information stage, (2) the product, the product is produced based on theoretical references in the first stage and based on consultation and discussion with supervisors, (3) field testing which includes product trials on students who have done real or virtual lab refutational learning.

The collected data was then analyzed quantitatively by using the Rasch Model in order to obtain conclusions that became a reference in improving the questionnaire. Winstaps explained that analysis based on equations in *Logit* raw data was converted into interval data (data "yes" and "no" were changed to 0 and 1). A *Logit* scale can express a person's skill level and item difficulty. However, in the case of non-test instruments, difficulty is defined as the tendency of respondents to find it difficult to give agreeable answers to certain items.

1. Respondents

The participants in this study were 53 people from high school in Makassar, South Sulawesi, Indonesia consisting of 12 males and 14 females. All of these samples have remedied misconceptions using a real (N=26) and virtual (N=27) refutational Laboratory. Data is collected using Google forms and then given to students accessed with the help of cellphones or laptops. The sample selection was carried out using the purposive sampling method which first provided a diagnostic test for misconceptions on several Physics concepts related to treatment. Furthermore, respondents were classified into three categories; scientific concepts, misconceptions, and lack of knowledge (Gurel et al., 2015). Respondents who are categorized as having misconceptions on physics concepts will then be used as responses in this study. The distribution of the number of samples in this study can be seen in table 1.

Table 1. The demographic profile of participants

Demography		Frequency	(%)
Gender	Boys	12	22,6
	Girls	41	77,4
Laborat-ory tipe	Real	26	49.0
	Virtual	27	51.0

2. Instrument

This instrument was developed with the aim of getting an overview related to the implementation of learning using the refutational laboratory method which was carried out with two approaches, namely real and virtual. 14 item questions were developed by assessing several aspects: assembling practical tools, reading the value of physical quantities by measuring instruments, reading value of physical quantities by measuring instruments, time required, challenge, and supporting for changing conception. Each of these aspects is further developed into several questions.

Efforts made in obtaining a questionnaire that can accurately measure what is intended are based on the following principles; (1) the questionnaire must be easy to answer and not lead to ambiguous interpretations. (2) all questions must use the same scale. (3) the questions and scales used must be based on preexisting questionnaires (Sumintono & Widhiarso, 2014).

Table 2. Instrumental aspects in exploring students' experiences related to the use of the refutational lab method

No	Aspect	Number of items	Adopted from	Total
1	Assembling practical tools	1	(Cuadros et al., 2021)	1
2	Reading the value of physical quantities by measuring instruments	2,3	(Cuadros et al., 2021)	2
3	Time required	4,5	(Garcia-Zubia et al., 2011)	2
4	Challenge	6,8	(Finstad, 2010; Bhansali et al., 2020)	2
5	Supporting for changing conception	7,9,10,11,12, 13,14	(Posner et al., 1982)	7

According to these characteristics, researchers developed the questionnaire using several aspects. In the motivational aspect, this questionnaire adopts instruments that have been developed by several previous researchers that have been used and have been declared valid, especially in the aspects of reading qualitative physics, data and measuring, efficiency, challenge, and motivation, while on the aspect of changing the author's conception of developing opinions (Posner et al., 1982) is related to general conditions related to changing the concept, namely: 1) raising dissatisfaction with existing concepts; 2) alternative concepts are easy to understand;3) alternative concepts make sense; and 4) are useful for additional explanations. These aspects were then developed into several questions to explore students' opinions related to the application of the refutational lab method in the CCLab model in practicum to remediate students' misconceptions.

3. Procedure, Data Analysis, and Rasch Measurement

The researcher conducted a conception test to ensure that students who experienced misconceptions were then used as samples. The test results are then used as a reference in the selection of samples. After that, the class is divided into virtual classes (respondents who will do practicum assisted by Phat applications) and real classes (respondents who will do real practicum in the laboratory) with three practical topics, namely hydro static pressure, pendulum oscillation period, and spring coefficient. At the end of all these stages students were given a questionnaire to assess their responses after being given treatment. The collected data were then analyzed using Winsteps software version 5.2.5.1 which was downloaded for trial and full version on the Google page. This application was used to analyze data using Rasch modeling. Analysis with this model has advantages that conventional analysis does not have. The flow of the research stages can be seen in figure 1.

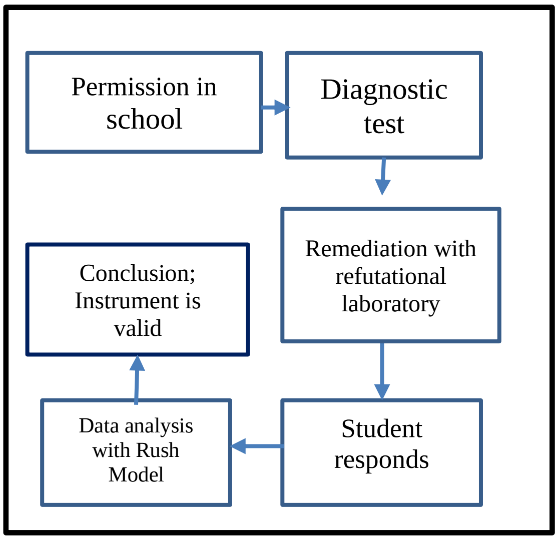


Figure 1. The research procedure

This instrument was analyzed quantitatively by taking into account the reliability and validity of the data. Reliability serves to determine the reliability of the instrument in measuring the variable to be measured even though it is measured by different researchers and respondents, while validity is a benchmark to find out the questionnaire functions well to measure what is being measured and is said to be feasible to use.

III. RESULTS AND DISCUSSION

This instrument was analyzed using the Rasch model to find out whether this instrument was feasible to use by measuring its validity and reliability after being tested on students. Before being analyzed using Winstaps software, the raw data is firstly converted into interval data. Tabel 3 shows the summery of data. It depicts the person mean measure with a logit value of 2.46. The person mean which is more than 0.0 shows that the tendency of respondents to respond more agreeably to positive statements and vice versa (Sumintono & Widhiarso, 2014).

Table 3. The summery of statistic base on person and items

	Person	Item
N	53	14
Mean	2.46	-0.25
SD	1.48	1.54
Mean Outfit MNSQ	1.05	1.05
Mean Outfit ZSTD	0.11	0.03
Separation	0.66	1.89
Cronbach's Alpha	0.61	
Reliability	0.31	0.82

The standard deviation of person is 1.48 and the mean measure is 2.46 logit which interprets that the person variation is quite good. The same thing is also seen in the items where the mean is -0.25 and the standard deviation is 1.54 indicating that the variety of data on the items is quite good as well. The mean Outfit mean-square and average outfit z-standardized (ZSTD) for persons and items show a range of 0.00 where these values are in the acceptance region. Meanwhile, the outfit mean-square (MNSQ) on the person and item is close to the value of 1.00 which shows that the closer to this number the better (Sumintono & Widhiarso, 2014). The grouping of persons and items is known based on the indicator of the separation value where the greater the value of separation, the quality of the instrument in this case the quality of respondents and items is getting better. Table 3 shows the separation for items of 0.66 which can be rounded up to one which shows that the data is not diverse, while the separation item shows a value of 1.89 which indicates the type of level and level of difficulty.

The reliability of an instrument shows how far repeated measurements will produce the same information using Cronbach's alpha value which provides an overview of the whole item. The value of Cronbach's alpha on the entire test item is 0.61 which indicates that the consistency of the instrument in measuring is sufficient. The Winstaps application also presents the person and item reliability values, respectively 0.31 and 0.82 which shows that the consistency of respondents' answers is weak but the quality of the items in the instrument is quite good.

Tabel 4. Unidimensionality instrument

Aspect	Eigenvalue	Observed
Raw Variance Explained by measure	5.4048	29.4%
Raw Variance Explained by Person	2.0731	11.3%
Raw Variance Explained by Items	3.3317	18.1%
Unexplained variance in 1st contrast	2.5924	14.1%

Table 4 shows the unidimensionality of instrument which is explained by measures of 29.4 percent which has met the minimum required value. Meanwhile, the Unexplained variance in the 1st contrast is a variance that cannot be explained by the instrument of 14.1% which meets the ideal characteristics in which the tolerable limit should not be more than 15% in order that it can be said that this questionnaire instrument is valid and reliable and can be used to measure the indicators (Soeharto et al., 2019; Sumintono & Widhiarso, 2014).

1. Item Fit Oder

Item fit order is one approach in the Rasch model to determine the level of difficulty of the item's difficulty level by looking at the quality of the item's suitability with the model. Item fit explains whether the items function properly in measuring or not. According to Sumintono & Widhiarso, (2014) the characteristics used to check items that do not fit (misfit) are; (1) the Outfit mean Square (MNSQ) value received is 0.5<MNSQ<1.5, (2) the Outfit Z-Standard (ZSTD) value received is -2.0<ZSTD < +2.0, and (3) the value The tolerated measure correlation points are 0.4 < Pt Measure Corr < 0.85. The item interpretation is said to be "very appropriate" if the three criteria are met, while if one of the criteria is not fulfilled then the item is said to be "appropriate", but if only one item is fulfilled the item is said to be "less appropriate", and if none of the three items are fulfilled then item is said to be "incompatible". Meanwhile, in determining the difficulty level of an item using the following characteristics: Difficult (M>+ 1SD), Medium (1SD≥M≥-1SD), and Easy M<-1SD) (Barma et al., 2015; Idris et al., 2021; Soeharto, 2021; Sumintono & Widhiarso, 2014).

Table 5. Processing results item quality and difficulty level

Number Item	*Outfit* MNSQ	*Outfit* ZFTD	PT Measure Corr	Interpretation	Measure	SD	Level of difficulty
1	0.87	-0.50	0.52	Very Appropriate	1.01	1.54	Medium
2	0.69	-1.01	0.53	Very Appropriate	0.59	1.54	Medium
3	0.98	-0.04	0.54	Appropriate	1.39	1.54	Medium
4	0.68	-0.72	0.46	Very Appropriate	0.10	1.54	Medium
5	1.07	0.35	0.61	Very Appropriate	1.81	1.54	Difficult
6	2.46	1.54	0.09	Inappropriate	-1.13	1.54	Medium
7	0.75	0.28	0.13	Appropriate	-2.33	1.54	Easy
8	0.99	0.09	0.48	Appropriate	0.44	1.54	Medium
9	0.66	-0.68	0.43	Very Appropriate	-0.08	1.54	Medium
10	2.08	1.43	0.16	Inappropriate	-0.80	1.54	Medium
11	1.05	0.26	0.35	Appropriate	0.10	1.54	Medium
12	0.85	0.20	0.18	Appropriate	-1.59	1.54	Easy
13	0.50	-0.84	0.40	Appropriate	-0.52	1.54	Medium
14	Minimum Measure			-	-3.56	1.54	Easy

Table 5 shows that most of the items fall into the “appropriate” and “very appropriate”. However, there are two items that only fulfill one characteristic so they are classified as unsuitable items. while the last item (Q14) is not identified in the system or only reads "minimum measure" with the result that the three items need to be repaired first.

Meanwhile, in classifying the level of difficulty of the items, most of the items are in medium level (71.43%), easy level (7.32%) and difficult level (2.44%).

2. Person Ability

Person ability describes personal ability in answering instruments. This percent ability can also be used to measure the item's difficulty level. However, the level of difficulty referred to in this instrument is the tendency of respondents to find it difficult to agree on an item or the tendency of respondents to answer no to positive statements and answer yes to negative statements.

Based on figure 4, the right side of the wright map is the distribution of person data with a logit measure from 0.18 to 4.55 (mean = 2.46 SD = 1.48) in the distribution of the logit value, the respondent's person ability can be categorized into 4 categories based on the logit value of item ((LVI) (Sumintono et al., 2014); low 24% ($0.18<LVI<1.36$), moderate 30% ($1.37<ILV<2.53$), high 21% ($2.54<ILV<3.7$), and very high 24% ($3.71<ILV<4.88$).

Item person analysis was carried out to understand the interaction between items and persons using the Wright Map which can provide an overview of students' abilities (the tendency of students to agree with most items) on the left side and the level of difficulty (the tendency of items that are difficult to agree with students) on the right side.

Wright's map is a person item map that can provide a comparison of people with items simultaneously in measurement content with an interval scale (logit) and assess the interaction between items and people and provide an overview of each person's abilities. If the item's ability matches the person's ability, it means that the person has a 50% chance ($p=0.5$) to answer correctly according to the item's difficulty level. If the location of the item is above the person, it indicates that the probability of the person in answering the item correctly is less than 50% ($p < 0.5$). Meanwhile, if the location of the item is lower than the person, it indicates that the instrument must be corrected. In the figure 4, it can be seen that most of the person is located above the item which shows that the difficulty level of the questionnaire is low, this is because most of the respondents agree with the positive statement given and disagree with the negative statement. This illustrates that the role of the learning method used in remediating misconceptions is considered by most respondents to be quite effective in remediating misconceptions about the given concept.

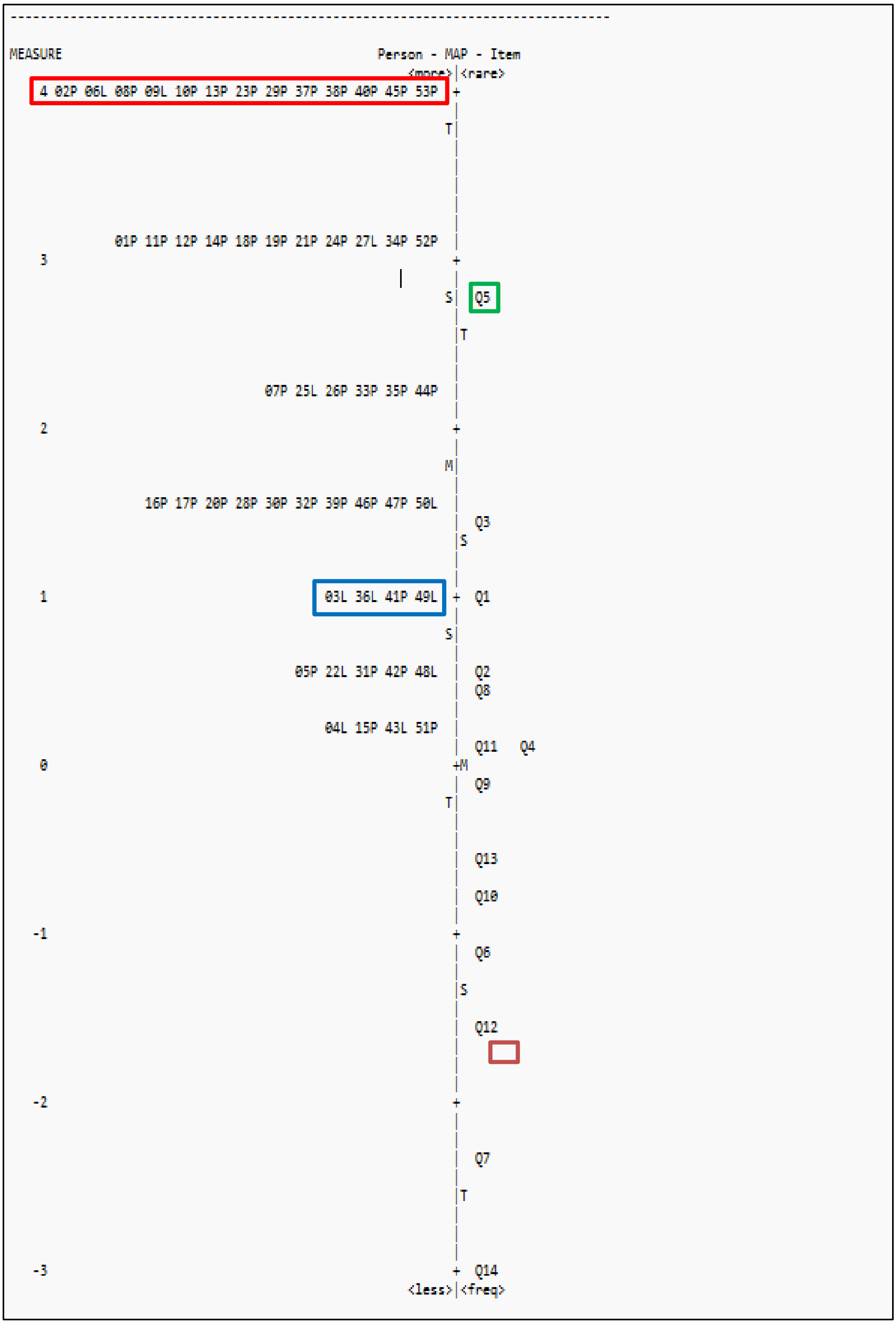
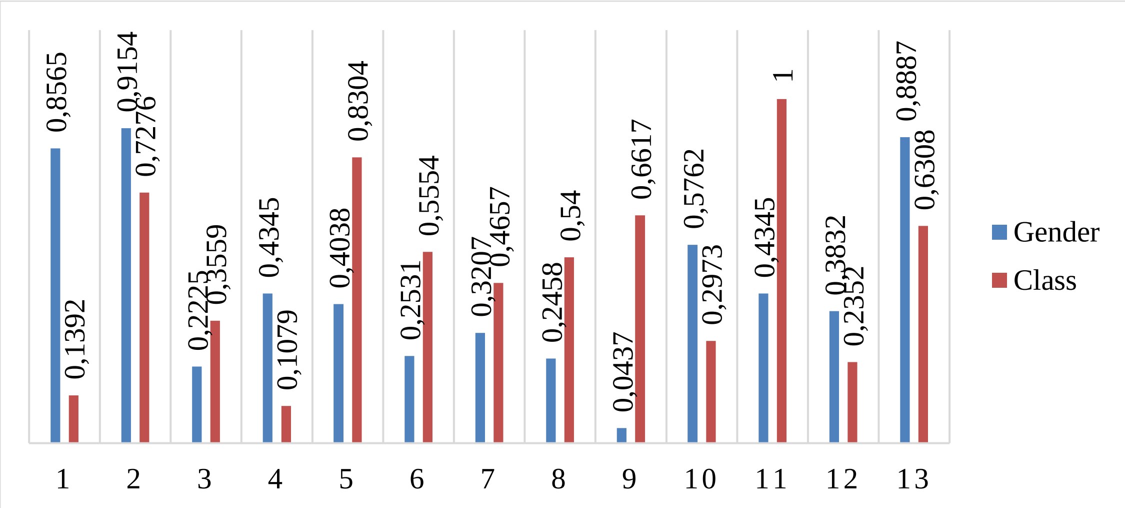


Figure 4. Wright person-item map of student perception

The picture also shows that in the red box there are 24% respondents (N=13) who most agree while the least agree 7% (N=4), this shows that the comparison of the most agree is much greater than the disagree. In terms of item distribution, the distribution is homogeneous, where Q5 (time required aspect) is the most difficult item. While Q14 is the easiest item with a logit of -3.00.

Differential Item Function (DIF) is an analysis in the item Rasch model that functions to detect bias in an item (Sumintono & Widhiarso, 2014). This analysis is needed to find out whether the items given have a bias in the respondent category or not. in this article the demographic grouping of gender groups and virtual and real class groupings. Bias on items can be identified based on the probability value of items where items that have a probability value of 5% (0.05) are grouped as biased items.



Fiqure 4. Analysis DIF item based on gender

a) DIF item base on gender

As explained above, the item bias indicator is a probability value <0.05 so the figure 4 shows in the blue color that item number 9 is classified as biased towards gender because the item probability value is 0.0435 so it can be said that the item is beneficial for one gender group but not for a group that is not. other.

b) DIF item based on Virtual and Real class

Grouping based on class is divided into Real class (treatment with real refutational laboratory learning which is carried out directly in the laboratory) and Virtual class (has been given treatment in the form of refutational laboratory learning assisted by Phat Colorado media. Based on the table above, it can be seen that all items have a probability above 0.05 which shows that the whole of item cannot deal with different types of laboratories learning approaches.

The output of this questionnaire can be used as a reference by researchers to find out students' views related to experiences after participating in learning using real and virtual RF methods. This instrument set can also be used by teachers in carrying out their work to evaluate the learning process carried out in order to improve the quality of their learning (Cuadros et al., 2021; Lutasari & Kartowagiran, 2019).

IV. CONCLUSION AND SUGGESTION

The questionnaire instrument was successfully developed to evaluate students' conceptions related to learning using the refutational laboratory method which was carried out with two types of treatment, namely real and virtual refutational laboratories. Based on the analysis of the reliability test using the Rasch model, the Cronbach's Alpha value, which measures how individuals interact with the item as a whole, is in the sufficient category. The value of item reliability is included in the good category, while person reliability provides an illustration that the consistency of students answers is classified as weak. Unidimentionality test using raw variance by measure shows that the instrument meets the minimum characteristics so that it is said to be valid and reliable in measuring what is being measured.

The item fit order analysis to obtain an overview of the quality and level of difficulty of the items revealed that most of the items had medium difficulty sticks and the quality of the items were mostly appropriate, but there were three items that only met one of the three required characteristics so that the total number of appropriate items was 11. To find out if there are items that are biased, a Differential Item Function test is carried out based on gender and type of treatment. The result is that there is one item that is biased by gender but in the other categories there is no bias at all.

ACKNOWLEDGMENTS

Sincere Thanks to Universitas Pendidikan Indonesia for providing us with a grant to carry out this research, as well as to my supervisor and all of people who have contributed to this research.

REFERENCES

Barma, S., Daniel, S., Bacon, N., Gingras, M.-A., & Fortin, M. (2015). Observation and analysis of a classroom teaching and learning practice based on augmented reality and serious games on mobile platforms. *International Journal of Serious Games*, *2*(2), 69-88. https://doi.org/10.17083/ijsg.v2i2.66

Bhansali, A., Angstmann, E., & Sharma, M. D. (2020). Aeq-physics : A valid and reliable tool to measure emotions in physics. *Proceedings of The Australian Conference on Science and Mathematics Education*, 93-98.

Cuadros, J., Serrano, V., Garcia-Zubia, J., & Hernandez-Jayo, U. (2021). Design and Evaluation of a User Experience Questionnaire for Remote Labs. *IEEE Access*, *9*, 50222–50230. https://doi.org/10.1109/access.2021.3069559

Cui, J., & Yu, S. (2019). Fostering deeper learning in a flipped classroom: Effects of knowledge graphs versus concept maps. *British Journal of Educational Technology*, *50*(5), 2308–2328. https://doi.org/10.1111/bjet.12841

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, *22*(5), 323–327. https://doi.org/10.1016/j.intcom.2010.04.004

Garcia-Zubia, J., Orduña, P., Angulo, I., Hernandez, U., Dziabenko, O., Lopez-Ipiña, D., & Rodriguez-Gil, L. (2011). Application and user perceptions of using the WebLab-Deusto-PLD in technical education. *Conference 1st Global Online Laboratory Consortium Remote Laboratories Workshop*, 1-6. https://doi.org/10.1109/GOLC.2011.6086780

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

Harman, G., Cokelez, A., Dal, B., & Alper, U. (2016). Pre-service science teachers' views on laboratory applications in science education: The effect of a two-semester course. *Universal Journal of Educational Research*, *4*(1), 12–25. https://doi.org/10.13189/ujer.2016.040103

Haryono, H. E., Aini, K. N., Samsudin, A., & Siahaan, P. (2021). Comprehensive teaching materials based on cognitive conflict strategies to reduce misconception of calories for junior high school students. *Jurnal Pendidikan Fisika*, *9*(3), 221–230. https://doi.org/10.26618/jpf.v9i3.5224

Hodson, D. (1996). Laboratory work as scientific method: Three decades of confusion and distortion. *Journal of Curriculum Studies*, *28*(2), 115–135. https://doi.org/10.1080/0022027980280201

Idris, M., Izmaimusah, D., & Ahmad. (2021). Instrument development of mathematics learning outcomes by the rasch model in elementary school to support the implementation of the 2013 curriculum. *Jisae: Journal of Indonesian Student Assesment and Evaluation*, *7*(2), 88–102. https://doi.org/10.21009/jisae.v7i2.21830

Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology*, *59*(3), 268–298. https://doi.org/10.1027/1901-2276.59.3.268

Lem, S., Baert, K., Ceulemans, E., Onghena, P., Verschaffel, L., & Dooren, W. V. (2017). Refutational text and multiple external representations as a method to remediate the misinterpretation of box plots. *Educational Psychology*, *37*(10), 1281–1300. https://doi.org/10.1080/01443410.2017.1333574

Liu, G., & Fang, N. (2016). Student misconceptions about force and acceleration in physics and engineering mechanics education. *Internasional Journal of Engineering Education*, *32*(1), 19–29.

Lutasari, S., & Kartowagiran, B. (2019). Developing instruments for student performance assessment in physics practicum: a case study of state senior high school of Magelang. *International Online Journal of Education & Teaching*, *6*(1), 104–114.

Marton, F., & Pang, M. F. (2008). *The idea of phenomenography and the pedagogy of conceptual change. In International handbook of research on conceptual change*. Routledge.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science education*, *66*(2), 211-227.

https://doi.org/10.1002/sce.3730660207

Rasmitadila., Aliyyah, R. R., Rachmadtullah, R., Samsuddin, A., Syaodih, E., Nurtanto, M., & Tambunan, A. R. S. (2020). The Perceptions of primary school teachers of online learning during the covid-19 pandemic period : A case study in Indonesia. *Journal of Ethnic and Cultural Studies*, *7*(2), 90–109. https://doi.org/10.29333/ejecs/388

Saputra, I. G. P. E., Sejati, A. E., & Nurazmi, N. (2021). Development of virtual laboratory system using EWB and zoom cloud in dynamic electricity practicum as a learning solution in the covid-19 pandemic. *Jurnal Pendidikan Fisika*, *9*(3), 262–272. https://doi.org/10.26618/jpf.v9i3.6066

Setiyoaji, W. T., Supriana, E., Latifah, E., Purwaningsih, E., & Praptama, S. S. (2021). The effect of learning simulation media on the students' critical thinking skills in vocational school during online learning in the covid-19 pandemic. *Jurnal Pendidikan Fisika*, *9*(3), 243–252. https://doi.org/10.26618/jpf.v9i3.5681

Sinatra, G. M. (2022). Motivational and emotional impacts on public (mis) understanding of science. *Educational Psychologist*, *57*(1), 1–10. https://doi.org/10.1080/00461520.2021.1975121

Sintema, E. J. (2020). Effect of COVID-19 on the performance of grade 12 students: Implications for STEM education. *Eurasia Journal of Mathematics, Science and Technology Education*, *16*(7), 1–6. https://doi.org/10.29333/EJMSTE/7893

Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 247–266. https://doi.org/10.15294/jpii.v8i2.18649

Soeharto, S. (2021). Development of a diagnostic assessment test to evaluate science misconceptions in terms of school grades : A Rasch measurement approach. *Journal of Turkish Science Education*, *18*(3), 351–370. Doi: 10.36681/tused.2021.78

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi)*. Trim Komunikata Publishing House.

Tabatabaee-Yazdi, M., Motallebzadeh, K., Ashraf, H., & Baghaei, P. (2018). Development and validation of a teacher success questionnaire using the rasch model. *International Journal of Instruction*, *11*(2), 129–144. https://doi.org/10.12973/iji.2018.11210a

Tumanggor, A. M. R., Supahar, S., Ringo, E. S., & Harliadi, M. D. (2020). Detecting students' misconception in simple harmonic motion concepts using four-tier diagnostic test instruments. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, *9*(1), 21–31. https://doi.org/10.24042/jipfalbiruni.v9i1.4571

Yunitasari, R., & Hanifah, U. (2020). Pengaruh pembelajaran daring terhadap minat belajar siswa pada masa covid 19. *Edukatif : Jurnal Ilmu Pendidikan*, *2*(3), 232–243. https://doi.org/10.31004/edukatif.v2i3.142