Received : 13 July 2023
Revised : 23 November 2023
Accepted : 24 December 2023
Online : 30 December 2023
Published: 31 December 2023

DOI: doi.org/10.21009/1.09201

Development and Validation of the Force Diagrams Representation Test

Judyanto Sirait[a)], Patrick Ricky, Hamdani

Department of Physics Education, Faculty of Teacher Training and Education, Universitas Tanjungpura, Pontianak, Indonesia

✉: [a)]judyanto.sirait@fkip.untan.ac.id

Abstract

Experts and students frequently use force diagrams as a physics representation to understand the force concept. This study aims to develop and validate a force diagram representation test that physics teachers in senior high schools can use. Research and development is implemented in this study by following some steps: analyze, define, design, implement, and evaluate. The 225 senior high school students in Pontianak are involved in this study, from the pilot study to the testing on a large scale. Data analysis includes measuring content validity, reliability index, and parameters of the test using the Rasch model. The test consists of 25 items in a multiple-choice format that cover three contexts: horizontal plane, inclined plane, and hanging object, and covers three situations, including the object being at rest, moving with constant velocity, and moving with constant acceleration. The results show that the content validity test obtained a score of 0.95 for the very good category, and the readability index results obtained a score of 7.42, which means it is easily understood by high school students. Rasch analysis shows that a unidimensional percentage is 30.5%, which means the test is able to measure one ability, and item reliability is 0.98. All items fit the Rasch model and had varying item difficulties. This study indicates that the force diagram representation test (FDRT) can be used by teachers as one of the instruments to measure students' understanding of force.

Keywords: force diagrams, representations, Rasch analysis

INTRODUCTION

In physics, representations such as text, pictures, tables, graphs, diagrams, equations and others (Bollen et al. 2017) are generally used to visualize concepts. Someone who can interpret representations and is able to transform one form of representation to another form is to have the ability of representation or multi-representation (Klein et al. 2017). This ability is very useful in learning physics because it can help students to understand physics concepts (Scheid et al. 2019), help to solve problems (Hamdani et al. 2019), and help students to get meaning from learning (Adawiyah & Istiyono 2021). As such, the representation ability constitutes as one of the important abilities that must be possessed by someone who is studying physics.

The concept of force is one of the physics topics studied by the students studying at the high school level. The application of force concept is often encountered in everyday life and can be understood through several forms of representation, one of which is force diagrams. The force diagrams as one of the diagrammatic representations are studied by students for the reason that it is easier to use in understanding the physical situation than abstract problems (Opfermann et al. 2017). Furthermore, force diagrams are used in problem solving because it helps describing the forces involved as well as predicting the motion of an object (Garcia-Lladó and López 2020). Next, the force diagrams are used as a link to further the finishing steps such as the use of mathematical equations to find the unknown quantities (Vignal and Wilcox 2022). Therefore, the techniques in describing force diagrams need to be learned by students in order to be able to well solve the questions about force (Linuwih et al. 2020).

The effectiveness of using force diagrams in solving questions about force has been performed by Rosengrant et al. (2009). The research involved two groups of students, namely the group that emphasizes the use of force diagrams (experimental class) and the group that places less emphasis on the use of style diagrams (control class). The research results showed that students from the experimental class spontaneously used force diagrams when solving multiple choice questions and getting better results if compared to students from the control class. This proves that the problems originally represented in the type diagram will be easily solved in terms of problem solving (Maries and Singh 2018). Therefore, the use of force diagrams really helps students in solving various questions.

In improving the diagram representation ability, there are various learning activities that can be applied by educators such as multi-representation based learning (Sirait and Silitonga 2016) which focuses on visualizing the physics concepts with different representations. Inquiry learning is implemented to train students' critical thinking with representations (Amanati et al. 2020). Science, Technology, Engineering, and Mathematics (STEM) apprroach develop students' represenations kills in learning Newton's laws (Mulyana et al. 2018). Besides, it is also necessary for the educators to pay attention to the tests used in investigating the force diagram representation ability. This is so as the test used in the evaluation of learning outcomes must be able to measure the concept being measured so that educators know students' understanding on the concept being tested (Adom et al. 2020). Several studies have been carried out on the test of force diagram representation by previous researchers. The research by Aviani et al. (2015) had developed the Free Body Diagram Test (FBDT) to explore students' understanding of force diagrams. It was a two-tier format that tested several contexts such as a rough plane, an inclined rough plane, mathematical swing, and circular motion. However, the number of questions was still limited as it did not include the context in which the object is connected to the pulley. Other research from Sirait et al. (2023) developed the Force Representational Competence Test (FRCT) to explore students' understanding of the concept of force through the use of representation force diagrams, mathematical representations, and verbal representations. The context tested in this test consists of a rough plane and a rough inclined plane and is assigned to college students. Then, this research is still in the pilot study stage so it cannot ensure as yet of good quality items. This is seen from several questions that have a low value of discrimination and a high level of difficulty indexes so that it requires further follow up.

Development of force diagram representation test especially for Senior High School still lacks its availability as a measuring tool for evaluating learning outcomes about such representation. Therefore, further development is needed, considering the importance of using force diagrams as a bridge in connecting with other representations (Sirait 2021). This study aims to develop and validate the force diagrammatic representation tests through a series of development procedures. The differences between this test and the previous test are the various contexts used, the number of students for data collection, and using Rasch model for analysis. The context varies such as horizontal plane, incline plane, hanging block, horizontal pulley, incline pulley, and Atwood machine are applied to this test. Then, the state of each context is divided into stationary and movable (such as constant velocity and acceleration). In developing this force diagram representation tests, the multiple- choice test format is used because it is easy to make an assessment, can be applied on a large number of students, and easy to find correct answers in a short time (Febriana 2019). The characteristic analysis in this study is also different from the previous research for the reason that there are some added parameters such as readability index analysis and Rasch models.

METHODS

This study involved 225 students (81 boys and 154 girls) from three high schools in the city of Pontianak. The determination of the sample in this study was carried out by means of purposive sampling to obtain samples based on the categories of high school, medium, and low. This is determined by considering the academic score of students according to the results of the National High School Physics Examination for the 2018/2019 Academic Year.

Research and Development (R&D) with the ADDIE development model (Analyze, Design, Develop, Implement, Evaluate) (Branch 2009) is used to develop Force Diagrams Representations Test (FDRT) adapted to the steps in preparing this test (Mardapi 2017). The steps for developing this test are shown in FIGURE 1.

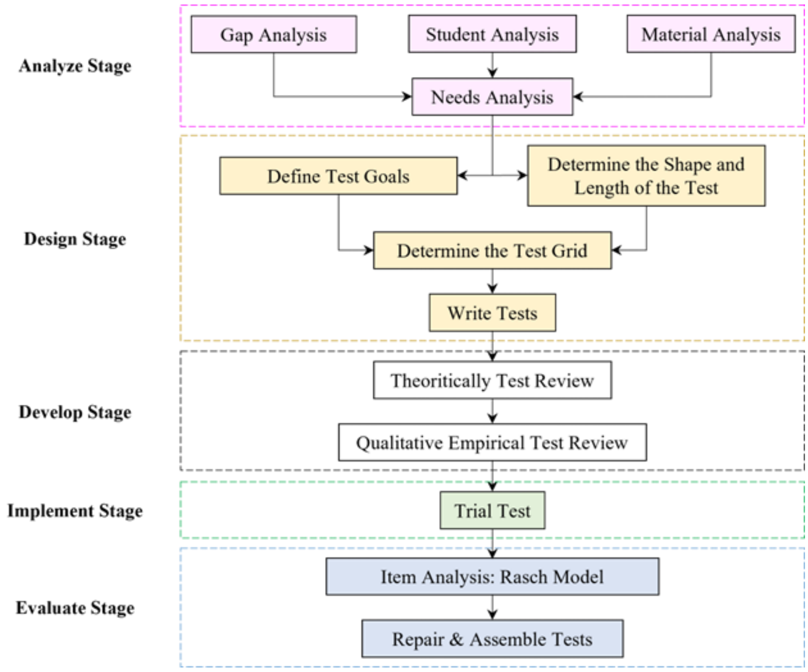

FIGURE 1. Development Procedures of FDRT

In the analysis stage, some data such as problem, student audience, documents and needs to consider are analyzed when designing the required tests. In the design stage, it defines several important factors in the initial test design, such as the objective of a measurement test, the format and test length, and the test grid. At the develop stage, it involves validating the tests, calculating readability metrics for the tests, and performing test studies. At the implement stage, tests are given to larger samples. At the evaluate stage, student feedback is analyzed using the Rasch model.

The data collection tools in this study used validation tables, questionnaires, and tests. The validation board is given to validators to receive suggestions to improve the developed test. In addition, the questionnaire is distributed to the students at the stage of testing the experimental qualitative test (experimental study) to obtain the candidate's answers to the given test. Finally, the test is used to elicit student responses, which can be used to analyze the item's characteristics across multiple parameters.

The analysis technique in this study consisted of content validity, readability index, and parameters of the Rasch model. First, the content validity test of a test is calculated using the Aiken Index which is mathematically formulated as:

$$V = \frac{\Sigma s}{n(c-1)}; \quad s = r - l_0 \tag{1}$$

where r is the score given by the validator, bing the lowest validity score, c is the highest validity score, and n is the number of validators. The interpretation of the value of content validity is presented in TABLE 1.

TABLE 1. Content Validity Interpretation

Value	Interpretation
0.81 – 1.00	Very high
0.61 – 0.80	High
0.41 – 0.60	Enough
0.21 – 0.40	Low
0.00 – 0.20	Very low

Secondly, the readability test index of the items is calculated using the formula from Smith (1961) which is mathematically formulated as:

$$RI = 1.56\overline{W}_{\mathrm{L}} + 0.19\overline{S}_{\mathrm{L}} - 6.49 \tag{2}$$

Where $\overline{W}_{\mathrm{L}}$ is the average word length and $\overline{S}_{\mathrm{L}}$ is the average sentence length. As for the index criteria legibility, it is divided into two, namely $RI < 6$ shows the text presented in the item questions can be understood by junior high school students while an RI score of ≥ 6 indicates the text presented on the items can be understood by senior high school students (Mahmuda 2011).

Finally, the Rasch model analysis of the characteristics of the items consists of several parameters including unidimensionality, reliability, item suitability level, wright map, and level item difficulty. These parameters are the last evaluation in the development of this test before being completely reassembled. Analysis of the Rasch model in this study was analyzed using Winstep version 5.1.4. The criteria for each parameter of the model Rasch is presented in TABLE 2.

TABLE 2a. Item Eligibility Criteria from the Rasch Model

Eligibility Aspects from the Rasch Model	Description	
Unidimensionality	Percentage of “raw variance explained by measure” > 20%.	
Person and Item Reliability	Weak	: < 0.67
	Pretty Good	: 0.67 – 0.80
	Good	: 0.81 – 0.90
	Very Good	: 0.91 – 0.94
	Special	: > 0.94
Wright Map	Describe the distribution of persons and items over the map.	
Item-fit	0.5 > MNSQ > 1.5, for Infit dan Outfit MNSQ	
	–2.0 > ZSTD > 2.0, for Infit and Outfit ZSTD	
Item-measure	Very difficult	: *measure logit > SD logit*
	Difficult	: 0 ≤ *measure logit ≤ SD logit*
	Easy	: *–SD logit ≤ measure logit* ≤ 0
	Very Easy	: *measure logit < –SD logit*

RESULTS AND DISCUSSION

Analyze Stage

The steps taken at this stage include: gap analysis, participant analysis students, material analysis, and needs analysis. Gap analysis is done to identify problems regarding the force diagram representation test. This analysis is done by collecting preliminary information from previous literature studies. The results of the literature study show that there are two tests which has been previously developed by previous researchers. Firstly, the Free Body Diagram Test (FBDT) developed by Aviani et al. (2015) has 12 items where the value of reliability is 0.78. Secondly, Sirait et al. (2023) developed the Force Representational Competence Test (FRCT) which has 30 items and the index of reliability is 0.80. Both FBDT and FRCT are intended for university level. Thus, there is still a lack of availability of representation tests force diagram as a measuring tool in evaluating learning outcomes at the high school level so that it still needs further development.

Student analysis was carried out to determine the sample of the testees in this research. Students who participated in this study were divided into pilot study groups and the main study group. The sample participating in this pilot study group was grade X students of a science class and the sample who participated in the main study group was 225 students (81 boys and 154 girls) from three different schools. The students involved in this study had studied the concept of force before.

A materials analysis was performed to identify the concepts used during test development. The concept of force constitutes the subject matter that grade X students of even semesters have studied and includes sub-topics such as Newton's laws, different forces, and their applications. However, understanding this concept requires an understanding of how to define a movement-oriented force of action. Thus, force diagrams play a role in conceptual understanding and problem solving. A needs analysis is performed to determine the type of product needed based on the results of gap analysis, student analysis, and literature analysis. On the basis of these three analyses, a specialized test is needed to measure the ability of high school students to express the force schema.

Design Stage

The design stage is the second phase of this research done to develop test specifications. The steps in this phase are: defining the test objective, defining the test format and duration, defining the test grid, and writing the test. This test is intended to assess the ability of high school students to express force diagrams. At that time, the test form used was a common multiple choice question because it was easy to evaluate, could be tested on a large number of students, and was easy to correct. The length of the test is designed to be 25 questions in 60 minutes. The indicators and test context are shown in TABLE 2. Then, examples of test items are presented in FIGURE 2. This question is about two blocks are connected with rope where the first block is on the frictionless surface and the other is hanging out on the rope. Students are asked the forces acting on the two blocks by selecting the appropriate of force diagrams.

TABLE 2b. Contexts and indicators of the FRDT

Context	Indicator	Item
Horizontal plane	Determine the appropriate force diagram for objects that stay on smooth and rough flat surface	1,5,6,7
	Determine the appropriate force diagram for objects that move at constant speed on a rough flat surface	8
	Determine the appropriate force diagram that move at constant acceleration on a smooth and rough flat surface	2,3,9
	Determine the appropriate force diagram for the system that moves in the context of horizontal pulley	19,20
Inclined plane	Determine the appropriate force diagram for objects the rest on a smooth and rough inclined plane	10,13
	Determine the appropriate force diagram for object that slide at a constant speed on a rough inclined plane	14
	Determine the appropriate force diagram for objects that move at constant acceleration on a smooth and rough inclined plane	11,12,15
	Determine the appropriate force diagram for the system which stationary and moving for the context inclined pullet	21,22,23
Hanging block	Determine the appropriate force diagram for objects that hanging and stationary	16
	Determine the appropriate force diagram for objects that move at constant acceleration subjected to the pull force	17,18
	Determine the appropriate force diagram for objects that moves in the context of the Atwood machine	24,25

Develop Stage

The development stage is the third phase of research done to validate the tests. The steps in this phase are to validate the test theoretically and validate the qualitative empirical test. Validating a theoretical test is done by validating the test content and calculating the readability index. Content validity is an objective parameter that ensures the accuracy of test equipment content and describes the overall learning content and learning objectives (Sudijono 2016). In other words, tests that can reveal the content of the concepts they are trying to measure have a high level of content validity (Febriana 2019). Content validation is performed in two phases. That is, the first phase was validated by his two expert verifiers of a physics education lecturer, and the second phase was validated by her three panel verifiers of his SMA physics teacher. This process was performed to obtain test improvement suggestions and scores from validators for analysis. Based on the validation results, the validator makes various suggestions for testing. First, an expert reviewer suggested providing a stylized vector arrow length description that has the same meaning as the style value, so the description was added to the question instructions. Additionally, the panel validator provides suggestions for replacing the word "situation" in the interrogative sentence. All of these are unfamiliar elements for high school students, so I will replace them with the word "conditions." In other proposals, the arrow component of the gravity vector has not been previously expressed, so we need to decompose the arrow component of the gravity vector in the inclined plane context. Participants are encouraged to identify the direction of object motion solely by the magnitude of the force vector arrow.

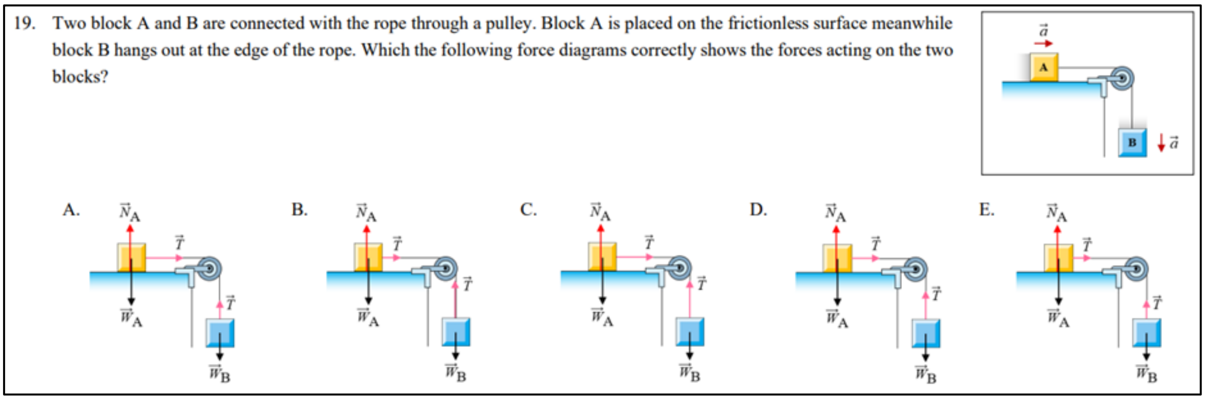
19. Two block A and B are connected with the rope through a pulley. Block A is placed on the frictionless surface meanwhile block B hangs out at the edge of the rope. Which the following force diagrams correctly shows the forces acting on the two blocks?

FIGURE 2. Example Of Item

Based on the validation results, the validator provides some suggestions regarding the test. First, the validator expert recommends presenting a description of the force vector arrow length with the same meaning as the force value so that the description will be added to the question instructions. In addition, the validator suggested replacing the word "situation" in the interrogative sentences of all items because high school students are not familiar, so the word "situation" should be replaced with the word "condition". Other suggestions are the need to construct gravity vector deviation components for inclined plane contexts since the design of gravity vector deviation component has not been presented before. Have students participate in determining the direction of motion of an object simply based on the size of the arrow of the force vector.

In addition to obtaining suggestions for improvement, the validator also evaluates validity test contents. The score obtained from the validator has been analyzed using an Aiken based index aspects of material, construction, and language. As for the calculation results for the material aspect, the Aiken index is obtained to have an average of 0.95 for the very good category. This score means that the material put on the test is in accordance with the basic competencies and subject matter, appropriate with question indicators, and are in accordance with the objectives of the measurement. Next, the calculation results of the construction aspect obtained an average Aiken index of 0.93 for the very good category. This score means that the questions have been formulated briefly and clearly, the images presented can be seen clearly, and have the most appropriate choice for answer. Then, the calculation results for the language aspect obtained an average Aiken index of 0.98 for the very good category. This score means that the item uses the correct grammar in accordance with the rules, communicative, and does not lead to multiple interpretations. Item 10 has the lowest index because the component of weight force is not drawn. Based on the three aspects, the overall Aiken average index is 0.95 for the very good category. Finally, the content validity value of each item is presented in FIGURE 3.

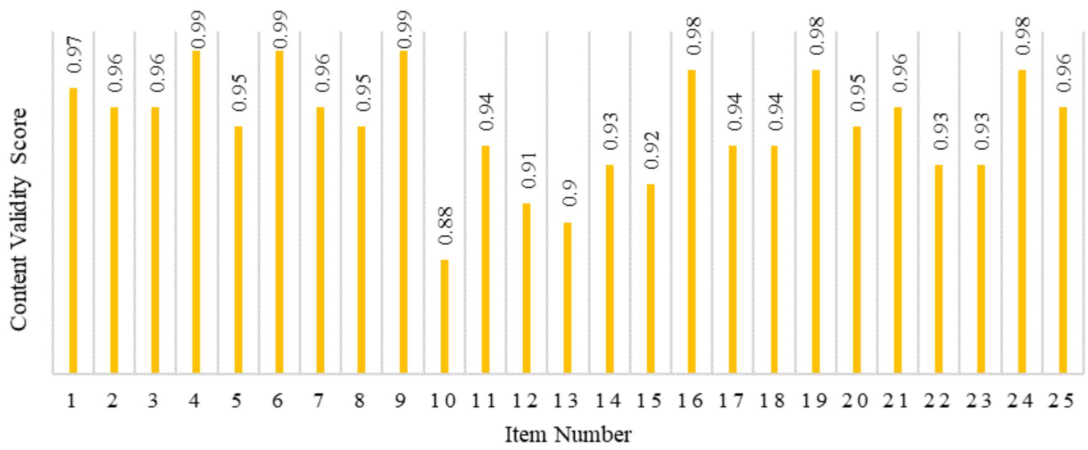

FIGURE 3. Content Validity Results

After the items are calculated for the validity value of their contents, then the items are calculated for their readability index. Readability index is a parameter that shows the ease of reading in understanding a text (Oyzon et al. 2015). The question text is an easy item understood by students depending on their level of education, namely RI < 6 for Junior High (SMP) and RI ≥ 6 for Senor High (SMA) level (Mahmuda 2011). The item readability index is calculated using the formula from Smith (1961). The calculation results show that the average readability index obtained at 7.42. Furthermore, the readability index results for each item are within the range of 6.03–8.93 as shown in FIGURE 4. This score means that all the items in the test have a readability index that is appropriate and easily understood by Senior High School students. Therefore, no elements are discarded or modified and all are to be used for further analysis.

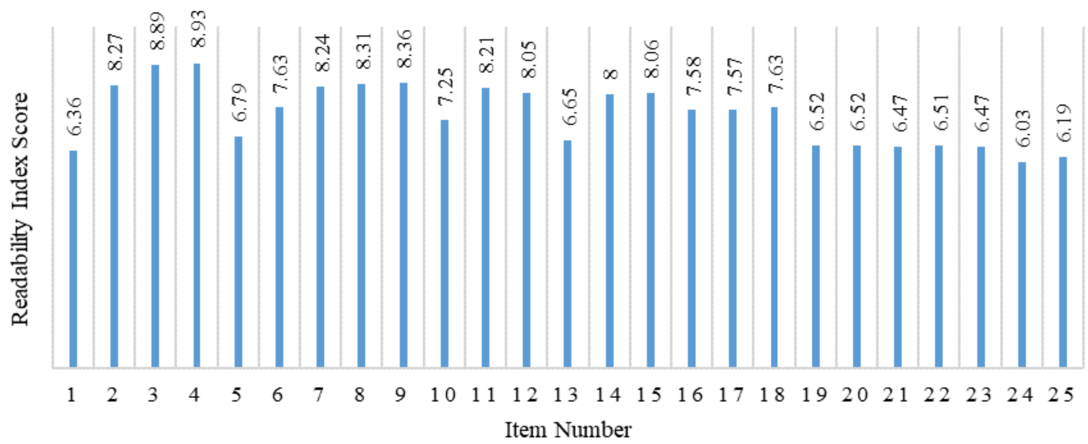

FIGURE 4. Readability Index Results

A qualitative empirical test validation was performed by conducting a pilot study. The purpose is to obtain the student's test answers through a questionnaire. Responses investigated in the pilot study included temporal, processing, conceptual, questioning, and stimulus aspects. The student's responses to the test indicate that the allotted time, 60 minutes, is sufficient to answer all the questions. In addition, the question text instructions are clear, there are no terms to obscure the students, each question is easy to understand, and the suggestions (pictures and style charts) are clearly presented. Therefore, the force diagram representation test can be used for large-scale trials (main trials).

Implement Stage

The implementation stage is the fourth phase of this research and is carried out to extensively pilot previously validated tests. The test will be conducted offline to 225 students from 3 different in Pontianak. Students were given 60 minutes to solve the test. The results obtained in the trial test are data on the reactions of the students when working on the force diagram representation test.

Evaluate Stage

The evaluation phase is the final phase of the study and is aimed at characterizing the available elements based on the Rasch model. The type of data used in this study is dichotomy data analyzed using the Winstep program version 5.1.4. The parameters used in analyzing element properties include uni-dimensionality, reliability, suitability element, difficulty element and Wright map.

Uni-dimensionality indicates that the developed test can only measure one ability alone (Planinic et al. 2019). Uni-dimensionality has the same meaning as validity construct. The value of uni-dimensionality is shown by the raw variance explained by measures, namely a percentage of more than 20% is required (Darmana et al. 2021). Based on the results of uni-dimensionality from the Winstep program, the raw variance explained by measures is 30.5%. This matter indicates that the test instrument is valid only to measure a single ability, namely force diagram representation.

TABLE 3. Rasch analysis of FDRT

Item Number	Total Correct Answer	Item Measure	Item-Fit				PMC
			Infit MNSQ	Outfit MNSQ	Infit ZSTD	Outfit ZSTD	
1	181	–1.41	0.93	1.10	–0.68	0.63	0.35
2	203	–2.29	0.93	1.33	–0.34	1.15	0.26
3	80	0.93	0.87	0.83	–2.03	–2.03	0.52
4	83	0.86	0.94	0.91	–1.04	–1.08	0.46
5	64	1.32	0.79	0.77	–3.01	–2.20	0.58
6	147	–0.52	0.97	0.91	–0.56	–0.87	0.41
7	63	1.34	0.83	0.83	–2.38	–1.56	0.54
8	128	–0.10	1.06	1.11	1.14	1.40	0.32
9	124	–0.02	0.90	0.85	–2.03	–2.06	0.50
10	201	–2.19	1.01	1.07	0.09	0.34	0.22
11	208	–2.59	0.94	0.59	–0.21	–1.33	0.31
12	184	–1.51	0.98	0.88	–0.14	–0.61	0.33
13	142	–0.43	1.14	1.46	2.35	4.46	0.18
14	105	0.38	1.09	1.09	1.66	1.26	0.31
15	163	–0.91	1.05	1.22	0.71	1.70	0.26
16	94	0.62	0.90	0.87	–1.85	–1.80	0.50
17	83	0.86	0.96	0.96	–0.64	–0.39	0.43
18	128	–0.02	1.13	1.12	2.38	1.55	0.27
19	81	0.91	1,07	1.11	1.13	1.26	0.32
20	125	–0.04	1.24	1.41	4.29	4.92	0.13
21	133	–0.21	0.97	0.94	–0.55	–0.71	0.42
22	65	1.29	1.04	1.12	0.61	1.08	0.33
23	48	1.76	0.96	0.86	–0.38	–0.95	0.41
24	82	0.89	1.06	1.13	0.91	1.52	0.33
25	70	1.17	0.98	1.04	–0.28	0.39	0.39

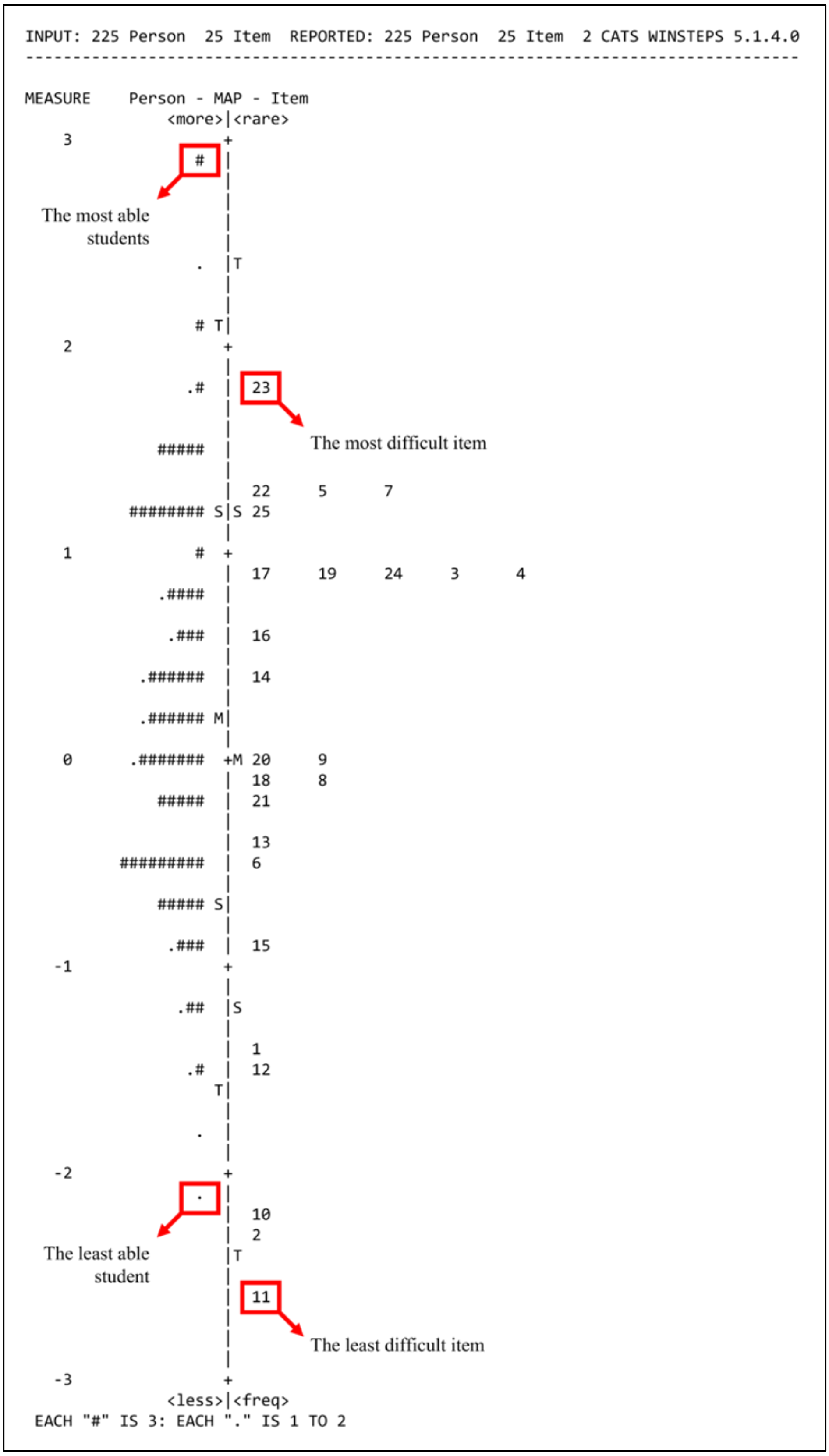

FIGURE 5. Wright Map

Reliability indicates that the developed test can provide consistent measurement results (Febriana 2019). Reliability in the Rasch model is divided into person reliability and reliability items. Person reliability is analogous to Cronbach's Alpha, which consistency shows the test takers when given test instruments that measure the same construct, while item reliability shows the consistency of the items being measured (Planinic et al. 2019; Bond et al. 2021). From the reliability results of the Winstep program, the person reliability of 0.70 is in the fairly good category, and the item reliability of 0.98 is in the special category. This shows that the consistency of students in answering the test is sufficient well and the quality of the items in the test can measure the ability of the diagrammatic representation of the students very consistently.

The suitability level of the items (item-fit) indicates whether the items can perform well during the measurement (Sumintono and Widhiarso 2015). The suitability of the items in the Rasch model determined by the MNSQ Infit and Outfit and ZSTD Infit and Outfit scores (Bond et al. 2021). Based on the item-fit results in TABLE 3, it shows that some of the items are in accordance with the Rasch model, except for item numbers 3, 5, 7, and 9 which have a ZSTD value below -2 as well as item numbers 13, 18, and 20 which have a ZSTD value above 2. However, in the case of item-fit in particular, the ZSTD value can be more than the limit because ZTSD is very sensitive to the sample amount so that it can be tolerated if the MNSQ value indicates a match against the measurement (Susac et al. 2018; Amelia 2021; Cvenic et al. 2022). Based on TABLE 3, item number 3, 5, 7, 9, 13, 18, and 20 have reasonable MNSQ Infit and Outfit values based on their provisions in range 0.5-1.5 so that it is considered suitable in measurement and in accordance with the Rasch model. So from that, all items are said to be appropriate and no items are revised or removed.

The Wright map shows the distribution of students' abilities and the distribution of levels difficulty of the items (Sirait 2023). The Wright map on the right shows the distribution of abilities students to the representation of force diagrams, where the distribution is based on value logit from person measure. Based on FIGURE 5, it was found that 119 samples (52.8%) were having sufficient understanding of the force diagram because the logit measure value is larger than zero, and found that 106 samples (47.1%) had a low understanding of diagrams force because the logit measure value is smaller than zero. Furthermore, three people were tagged with "#" has a logit value of 2.92 indicating that they had the highest ability where they could almost answer all the questions correctly, and one person who was marked by "." having a logit value of –2.12 indicates that he had the lowest ability but could answer some questions correctly.

The Wright map on the right shows the distribution of the item difficulty levels, where the distribution is based on the logit value of the measured item. The level of difficulty of the items (items-measure) shows the ease and difficulty of the item based on the total score of the correct answers (Hamidah and Istiyono 2022). Based on the map, the items that were the most difficult to work on by students were indicated by item number 23 with a logit value of 1.76. This matter discusses the context of the inclined pulley for a beam that hangs downwards and a block on an inclined plane moves up to the top. The logical reason most students got this question wrong is that they didn't pay attention to carefully measuring the string tension arrows that were supposed to match. TABLE 3 also shows that only 48 of 225 students (21.3%) answered these questions. Additionally, the easiest task for students to tackle is identified by task number 11 with a logit value of -2.59. This item discusses a sliding block on a smooth inclined plane. Item number 11 can almost be answered by all students, namely as many as 208 of 225 students (92.4%).

The difficulty level of each item in detail can be observed in TABLE 3. A good test is characterized by items that are neither too easy nor too difficult for students (Febriana 2019). As a result of the item distribution, 4 items are in the very difficult category (numbers 5, 7, 22, 23), 8 items are in the difficult category (numbers 3, 4, 14, 16, 17, 19, 24, 25), 8 items (numbers 6, 8, 9, 13, 15, 18, 20, 21) are in the easy category and 5 items are in the very easy category (numbers 1, 2, 10, 11, 12). Based on the difficulty distribution of each item, it could be said that the items in this test have different difficulty levels. This means that there are not too many difficult or too easy questions. The evaluation results of the Rasch model as a whole show that the force diagram representation test (FDRT) has good item characteristics in terms of uni-dimensionality, reliability, the level of suitability and the level of difficulty of the items so that the test could be corrected and assembled into a complete test.

CONCLUSION

Data analysis indicated that the force diagram representation test (FRDT) was applicable. This is evidenced by the enormous value of content validity based on material aspect, structure and language.

Furthermore, the high readability index indicates that high school students can correctly grasp and process information from the text of each item. Finally, the results of the analysis of the Rasch model also shows good results where this test has sufficient strong constructs in measuring force diagram representation abilities, well-performing, consistently measurable, and varying difficulty components. This research can be further developed for tests that measure multiple presentation skills.

ACKNOWLEDGMENT

The authors would like to thank Universitas Tanjungpura for supporting the study in developing representations test by DIPA 2022. We also thank to validators and students who participated in this study.

REFERENCES

Adawiyah, R & Istiyono, E 2021, 'Assessment Instrument on Measuring Physics Verbal Representation Ability of Senior High School Students', *in Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)*, pp. 591-599.

Adom, D, Mensah, JA & Dake, DA 2020, 'Test, measurement, and evaluation: Understanding and use of the concepts in education', *International Journal of Evaluation and Research in Education*, vol. 9, no. 1, pp. 109-119.

Amanati, AY, Wasis & Ibrahim, M 2020, 'The Effectiveness of Learning Instrument of Multiple Representations-Based Inquiry Model to Train Critical Thinking Skills in Physics Lesson', *JPPS (Jurnal Penelitian Pendidikan Sains)*, vol. 9, no. 1, p. 1772.

Amelia, RN 2021, 'Identifikasi Item Fit dan Person Fit dalam Pengukuran Hasil Belajar Kimia', *Jurnal Ilmiah WUNY*, vol. 3, no. 1, pp. 13-26.

Aviani, I, Erceg, N & Mešić, V 2015, 'Drawing and using free body diagrams: Why it may be better not to decompose forces', *Physical Review Special Topics - Physics Education Research*, vol. 11, no. 2, pp. 1-14.

Bollen, L, Van Kampen, P, Baily, C, Kelly, M & De Cock, M 2017, 'Student difficulties regarding symbolic and graphical representations of vector fields', *Physical Review Physics Education Research*, vol. 13, no. 2, pp. 1-17.

Bond, T, Yan, Z & Heene, M 2021, 'Applying The Rasch Model: Fundamental Measurement in the Human Sciences 4th edn', Routledge: Taylor & Francis, New York.

Branch, RM 2009, 'Instructional Design: The ADDIE Approach', *Springer*, United States of America.

Darmana, A, Sutiani, A, Nasution, HA, Ismanisa, I & Nurhaswinda, N 2021, 'Analysis of Rasch Model for the Validation of Chemistry National Exam Instruments', *Jurnal Pendidikan Sains Indonesia*, vol. 9, no. 3, pp. 329-345.

Febriana, R 2019, 'Evaluasi Pembelajaran', Bumi Aksara, Jakarta.

Garcia-Lladó, À & López, V 2020, 'Beyond Recurring Free-Body Force Diagrams: Educational Pros and Cons of Alternative Means of Representing Forces and Interactions', *The Physics Teacher*, vol. 58, no. 7, pp. 504-508.

Hamdani, H, Mursyid, S & Sirait, J 2019, 'Using physics representation worksheet to enchance students' understanding and performance about force', *Journal of Physics: Conference Series*, vol. 1157, no. 3. pp. 1-6.

Hamidah, N & Istiyono, E 2022, 'The quality of test on National Examination of Natural science in the level of elementary school', *International Journal of Evaluation and Research in Education*, vol. 11, no. 2, pp. 604-616.

Klein, P, Müller, A & Kuhn, J 2017, ‘Assessment of representational competence in kinematics’, *Physical Review Physics Education Research*, vol. 13, no. 1, pp. 1-18.

Linuwih, S, Asih, P & Ellianawati 2020, ‘The Different Ability of Free Body Diagram (FBD) Representation in Newton’s Law Topic Based on Students’ Thinking Style’, *Journal of Physics: Conference Series*, vol. 1567, no. 2, pp. 1-6.

Mahmuda, D 2011, ‘Secondary Analysis Tentang Tes Skripsi-Skripsi Mahasiswa Pendidikan Fisika FKIP UNTAN Tahun 2007-2009 pada Materi Mekanika’, *Skripsi*, FKIP : Untan, Pontianak.

Mardapi, D 2017, ‘Pengukuran, Penilaian, dan Evaluasi Pendidikan 2nd edn’, Parama Publishing, Yogyakarta.

Maries, A & Singh, C 2018, ‘Case of two electrostatics problems: Can providing a diagram adversely impact introductory physics students’ problem solving performance?’, *Physical Review Physics Education Research*, vol. 14, no. 1, pp. 1-14.

Matejak Cvenic, K, Planinic, M, Susac, A, Ivanjek, L, Jelicic, K & Hopf, M 2022, ‘Development and validation of the Conceptual Survey on Wave Optics’, *Physical Review Physics Education Research*, vol. 18, no. 1, pp. 1-11.

Mulyana, KM, Abdurrahman & Rosidin, U 2018, ‘Implementasi Pendekatan Science, Technology, Engineering, and Mathematics (STEM) untuk Menumbuhkan Skill Multirepresentasi Siswa SMA pada Materi Hukum Newton Tentang Gerak’, *Jurnal Pendidikan Fisika*, vol. 7, no. 2, pp. 69-75.

Opfermann, M, Schmeck, A & Fischer, HE 2017, ‘Multiple Representations in Physics and Science Education - Why Should We Use Them?’, *Springer*, pp. 1-22.

Oyzon, VQ, B. Corrales, J & M. Estardo, JW 2015, ‘Validation Study of Waray Text Readability Instrument’, *International Journal of Evaluation and Research in Education (IJERE)*, vol. 4, no. 2, pp. 45-53.

Planinic, M, Boone, WJ, Susac, A & Ivanjek, L 2019, ‘Rasch analysis in physics education research: Why measurement matters’, *Physical Review Physics Education Research*, vol. 15, no. 2, pp. 1-14.

Rosengrant, D, Van Heuvelen, A & Etkina, E 2009, ‘Do students use and understand free-body diagrams?’, *Physical Review Special Topics - Physics Education Research*, vol. 5, no. 1, pp. 1-13.

Scheid, J, Müller, A, Hettmannsperger, R & Schnotz, W 2019, ‘Improving learners’ representational coherence ability with experiment-related representational activity tasks’, *Physical Review Physics Education Research*, vol. 15, no. 1, pp. 1-23.

Sirait, J & Silitonga, HTM 2016, ‘Representations Based Physics Instruction to Enhance Students’ Problem Solving’, *American Journal of Educational Research*, vol. 4, no. 1, pp. 1-4.

Sirait, J 2021, ‘Multirepresentasi dalam Penyelesaian Soal Fisika’, *Fahruna Bahagia Press*, Pontianak.

Sirait, J 2023, ‘Development and Validation of Physics Learning Motivation Survey (PLMS) using Rasch Analysis’, *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 5, pp. 4063-4069.

Sirait, J, Firdaus, F, Hidayatullah, MMS & Habellia, RC 2023, ‘Development and Validation of Force Test to Assess Physics Education Students’ Representational Competence’, *Jurnal Pendidikan Sains Indonesia*, vol. 11, no. 2, pp. 306-317

Smith, EA 1961, ‘Devereux Readability Index’, *Journal of Educational Research*, vol. 54, no. 8, pp. 298-303.

Sudijono, A 2016, ’Pengantar Evaluasi Pendidikan’, *Rajawali Pers*, Jakarta.

Sumintono, B & Widhiarso, W 2015, ’Aplikasi Pemodelan Rasch pada Assessment Pendidikan’, *Penerbit Trim Komunikata*, Cimahi.

Susac, A, Planinic, M, Klemencic, D & Milin Sipus, Z 2018, 'Using the Rasch model to analyze the test of understanding of vectors', *Physical Review Physics Education Research,* vol. 14, no. 2, pp. 1-6.

Vignal, M & Wilcox, BR 2022, 'Investigating unprompted and prompted diagrams generated by physics majors during problem solving', *Physical Review Physics Education Research,* vol. 18, no. 1, pp. 1-19.