JURNAL PEMBELAJARAN FISIKA
http://jurnal.fkip.unila.ac.id/index.php/JPF
VOL. 12 (1), JUNE 2024, 37 - 50 || P-ISSN 2302-0105, E-ISSN 2684-9828
RESEARCH ARTICLE

Rasch Analysis of the Force and Motion Conceptual Evaluation Test: Validity and Reliability in Measuring Force and Motion Understanding of Students

Nur Akhyar Basri, Ummi Salmah, Ivan Danar Aditya Irawan, Ristanti Mulia Indraloka, Parno*

Physics Department, Faculty of Mathematics and Natural Sciences, Malang State University, Indonesia

**Corresponding Author, E-mail: parno.fmipa@um.ac.id*

Article History

Received:
15 Jue 2024

Accepted:
22 July 2024

Published:
28 July 2024

Keywords

Concept understanding, force and motion, high school students, instrument test, Rasch model

Abstract

Understanding Force and Motion is fundamental in physics education as it forms the basis for more complex concepts and has practical applications in various fields, from engineering to everyday problem-solving. This paper aims at assessing the quality of test questions using the Rasch model to gauge students' understanding of Force and Motion within physics education. The significance of accurately assessing these concepts cannot be overstated, as it ensures that students have a solid foundation for future learning. Adopting a descriptive qualitative approach, the research employed the FMCE (Force and Motion Conceptual Evaluation) test instrument alongside Rasch modelling. The study involved 35 high school students who had covered the Force and Motion curriculum. Analysis with Winstep software (Version 3.65.0) revealed that items 1 and 7 were invalid. The instrument demonstrated commendable reliability, with an item reliability of 0.73. Difficulty level analysis identified five questions as outliers, categorised as either very difficult or very easy. The discrimination analysis confirmed that the instrument effectively differentiated between students who answered correctly and those who did not. Overall, the FMCE exhibited solid validity and high reliability, although some items necessitate revision. The study's limitations, particularly the small sample size, may affect the generalisability of the findings. Despite these limitations, the study provides valuable insights into the assessment of Force and Motion concepts in high school students, though caution is advised when interpreting the results. Future research should consider a larger sample size and diverse educational contexts to enhance the robustness and applicability of the findings.

DOI: http://dx.doi.org/10.23960/jpf.v12.n1.202404

INTRODUCTION

One of the aims of physics education is for students to grasp physics concepts and apply them in everyday life. Conceptual understanding denotes the ability to utilise knowledge in different contexts from those previously learned. As per Balka et al. (2021), conceptual understanding is the ability of students to think within a particular context, necessitating the correct application of concepts, descriptions, relationships, or representations. In physics, conceptual understanding also encompasses the ability to integrate various knowledge to solve problems by applying familiar concepts in new situations (Hake & Gibbons, 2020; Browne & Schreiber, 2020; Ding & Beichner, 2021). The objective of physics learning is for students to thoroughly comprehend physics concepts and be capable of applying them to solve everyday issues (Didik et al., 2020; Mestre et al., 2011; Ryan et al., 2016). Learning success is assessed by students' ability to employ physics concepts to address daily life problems (Ismunandar et al., 2022).

Some studies provide an in-depth examination of prevalent misconceptions in physics education, specifically concentrating on concepts such as motion and force. McDermott and Redish (2021) highlight the critical need for developing improved instructional strategies to address these misconceptions and enhance students' conceptual understanding. Building on this, Czerniak and Johnson (2020) explore persistent misconceptions among high school students regarding motion and force, discussing their implications for physics education and suggesting corrective approaches. Zhu and Dong (2019) further review misconceptions across various educational levels, from elementary to high school, providing insights into how these issues can be effectively tackled. Singh and Chabay (2022) delve into strategies for improving conceptual understanding specifically related to motion and force, proposing methods to overcome these common misunderstandings. Finally, Finkelstein and Leong (2021) emphasise the need for ongoing education and refined instructional practices to address specific conceptual difficulties and misconceptions among high school students, reinforcing the importance of continuous improvement in teaching methods.

Field observations reveal that concept understanding tests on Motion and Force materials often lack valid and reliable instruments. Interviews with high school physics teachers indicate that daily assessments consist of questions created by the teachers themselves, without undergoing validity and reliability tests. Teachers tend to reuse practice questions previously given during lessons. Consequently, most of the instruments employed have not yielded optimal results (Kadir Masalesi, 2022; Misbah et al., 2022; Nehru et al., 2022; Yusuf et al., 2022).

To evaluate conceptual understanding, an assessment is required that can provide insights into the aspects and components of learning that need improvement (Miller & Kopp, 2020; Wang & Wang, 2019; Kali & Ronen, 2019; Hattie & Timperley, 2020). This assessment is crucial for measuring student achievement on the specific characteristics being assessed. Enhancing the quality of measurement instruments can yield more accurate and reliable results in gauging student understanding (Wang & Black & Wiliam, 2020; Wang, 2021). Therefore, a suitable research instrument is needed to assess conceptual understanding. The Rasch model offers a solution to the limitations of classical statistical models in analysing an instrument (Medriati et al., 2022).

The Rasch Model is a probabilistic framework that elucidates the interaction between individuals and test items through two parameters: item difficulty and individual ability (Planinic et al., 2019). Rasch model analysis can identify item bias and ascertain correctly answered questions, ensuring that students are not disadvantaged by inaccurate measurements (Sulman et al., 2021). This model effectively creates an accurate, linear, and objective measurement scale, offering a thorough analysis of students' conceptual understanding abilities (Bond & Fox, 2019; Meyer & Timm, 2021). The Rasch model can pinpoint interactions between items and students on the same linear and interval scale, utilising logit unit values to depict the probability of correct answers (Bond & Fox, 2019). Research demonstrates that measurement instruments can discern the relationship between student ability and item difficulty (Palimbong et al., 2019). Another capability of the Rasch Model, when utilised with the Winsteps program, is its ability to detect inconsistencies in individual response patterns. These inconsistencies, or mismatched answers, can be identified by comparing individual responses to the ideal model based on their ability. This feature can aid teachers in determining the consistency of students' thinking and in identifying any instances of cheating by students.

Various studies have effectively utilised the Rasch Model to analyse physics education, particularly regarding motion and force materials, there are several limitations to consider. Firstly, the Rasch Model's reliance on precise item and person calibration may not account for all sources of variability in student responses, such as contextual or motivational factors (McDermott & Redish, 2021). Additionally, while the model provides a structured approach to understanding student ability and item difficulty, it may not fully capture the complexity of students' misconceptions or the subtleties of instructional effectiveness (Singh & Chabay, 2022). Furthermore, the generalisability of findings from studies employing the Rasch Model can be limited by sample size and demographic variability, which may affect the robustness of the conclusions drawn about instructional materials and teaching strategies (Zhu & Dong, 2019). Lastly, the focus on quantitative metrics provided by the Rasch Model may overlook qualitative aspects of student learning and engagement, which are also crucial for a comprehensive understanding of educational outcomes (Czerniak & Johnson, 2020; Finkelstein & Leong, 2021). Therefore, this study aims to analyse the existing items to ensure that the instrument can accurately measure students' conceptual understanding. The objective is to evaluate the quality of items in assessing students' conceptual understanding by considering the validity, reliability, difficulty index, and differentiation index of each tested item. This has the potential to enhance learning evaluation and provide more precise information about students' abilities and the quality of the proposed questions, using the Rasch Model approach. The Rasch model, as described by Sumintono & Widhiarso (2015), is similar to the 1PL model in its focus on measuring difficulty.

METHOD

The research method employed in this study was descriptive qualitative research, adhering to procedures for organising and conducting observations. Researchers did not implement any special interventions to alter the students but rather observed the existing conditions using current instruments. This research focused on the analysis stage of the FMCE (Force and Motion Conceptual Evaluation) test instrument, which utilised Rasch modelling (Creswell, 2018) and criteria adopted from the research by Austvoll-Dahlgren et al. (2017) and Sumintono & Widhiarso (2015). The test subjects comprised 35 MAN 2 Kuantan Singingi students who had studied Motion and Force material. The test results were analysed using the Winsteps Version 3.65.0 program.

The stages of this research analysis include item validity testing, item reliability testing, problem difficulty level assessment, and distinguishing power analysis. The question instrument used in this study comprised a total of 15 items, which were analysed using the Rasch model. In the Rasch model, determining the validity of a question requires meeting specific criteria (Table 1). Suitable questions were identified, followed by an analysis and conclusion of students' conceptual understanding using the Winsteps Version 3.65.0 program, focusing on students' mastery of concepts, particularly in motion and force material.

Table 1. Person fit criteria levels

Person Fit Criteria	
Outfit dan Infit mean square (MNSQ) Values	0.5 < MNSQ < 1.5
Outfit dan infit Z-standar (ZSTD)	-2.00 < ZSTD < +2.0
Point Measure Correlation (Pt Mean Corr)	0.4 < Pt Measure Corr < 0.85

A reliability test is also utilised to evaluate the consistency and stability of the results produced by the test instrument. Table 2 outlines the criteria for the reliability coefficient according to the Rasch model. Besides that, item difficulty is categorised based on Measure logit and Standard Deviation (SD) logit values, and is divided into five categories as shown in Table 3.

Table 2. Reliability coefficient criteria

Rasch Model Reliability Coefficient	
Value Criteria	**Description**
>0,8	Very High
0,70 – 0,80	High
0,60 – 0,70	Moderate
0,50 – 0,60	Low
<0,50	Very Low

Table 3. Criteria for question item difficulty

Value Criteria	**Item Difficulty Category**
Measure logit < SD Logit	Very Easy (Outliners)
- SD Logit < Measure Logit < -0,54	Easy
- 0,54 ≤ Measure logit ≤ 0,54	Medium

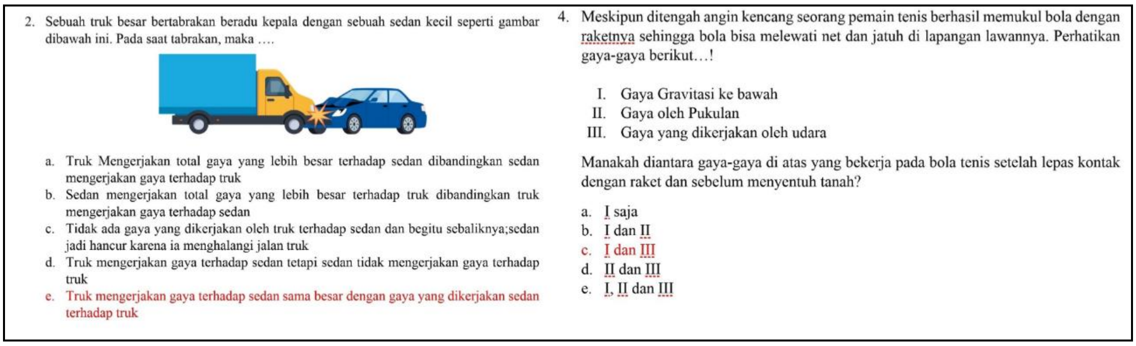
Publisher: Physics Education Study Program, Faculty of Teacher Training and Education, Universitas Lampung in Collaboration with the Physical Society of Indonesia (PSI)

0,54 ≤ measure logit ≤ SD logit	Difficult
Measure logit > SD logit	Very Difficult (Outliners)

The distinguishing power analysis is also used to differentiate between students of high and low ability. A high question differentiation index signifies that the question is more effective at distinguishing between varying levels of student ability. If the distinguishing power of the question is negative, it suggests that a greater number of student groups have not understood the material. This distinguishing power can be calculated using the following equation.

$$H = \frac{(4 \times separation) + 1}{3}.$$

Where H represents strata equation. The Strata (H) equation in the context of Rasch measurement is used to assess the separation of strata or groups of individuals in terms of their ability levels. Specifically, it is used to evaluate how well the measurement instrument can differentiate between different levels of ability among the respondents.



2. Sebuah truk besar bertabrakan beradu kepala dengan sebuah sedan kecil seperti gambar dibawah ini. Pada saat tabrakan, maka ....
   a. Truk Mengerjakan total gaya yang lebih besar terhadap sedan dibandingkan sedan mengerjakan gaya terhadap truk
   b. Sedan mengerjakan total gaya yang lebih besar terhadap truk dibandingkan truk mengerjakan gaya terhadap sedan
   c. Tidak ada gaya yang dikerjakan oleh truk terhadap sedan dan begitu sebaliknya;sedan jadi hancur karena ia menghalangi jalan truk
   d. Truk mengerjakan gaya terhadap sedan tetapi sedan tidak mengerjakan gaya terhadap truk
   e. Truk mengerjakan gaya terhadap sedan sama besar dengan gaya yang dikerjakan sedan terhadap truk

4. Meskipun ditengah angin kencang seorang pemain tenis berhasil memukul bola dengan raketnya sehingga bola bisa melewati net dan jatuh di lapangan lawannya. Perhatikan gaya-gaya berikut…!

   I. Gaya Gravitasi ke bawah
   II. Gaya oleh Pukulan
   III. Gaya yang dikerjakan oleh udara

   Manakah diantara gaya-gaya di atas yang bekerja pada bola tenis setelah lepas kontak dengan raket dan sebelum menyentuh tanah?

   a. I saja
   b. I dan II
   c. I dan III
   d. II dan III
   e. I, II dan III

Figure 1. Example of FMCE

RESULT AND DISCUSSION

As previously discussed, this study was conducted through several key stages: item validity testing, item reliability testing, assessment of problem difficulty levels, and distinguishing power analysis. These stages were applied to the FMCE (Force and Motion Conceptual Evaluation) test instrument to ensure a thorough evaluation of its effectiveness. One example of FMCE used is presented in Figure 1. The validity testing aimed to confirm the accuracy of each item in measuring the intended concepts, while reliability testing assessed the consistency of the results. Additionally, the analysis of problem difficulty levels offered insights into the varying degrees of challenge posed by the test items, and the distinguishing power analysis evaluated the test's ability to differentiate between students with varying levels of ability.

Based on Table 4, for Infit and Outfit MNSQ, the validity is deemed satisfactory if the Mean Square value falls within the range of 0.5 to 1.5. This range signifies that the questions align well with student responses and the Rasch model, with only a minimal discrepancy between actual and expected responses. Additionally, for Infit ZSTD, the validity is considered adequate if the Z-Standard value lies between -2.0 and 2.0. This range suggests that the disparity between student responses and the Rasch model is within statistically acceptable limits. Therefore, meeting these two criteria indicates that the questions have achieved the expected level of validity according to the Rasch model. Following the validity testing of each item using the Rasch model, the results of the analysis concerning the validity of the items related to the understanding of motion and force concepts can be observed in Table 5.

Table 4. Results of Item Analysis Using the Rasch Model

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
7	1	35	2.05	1.02	.95	.25	.40	-.20	.28	.12	97.1	97.1	S7
5	2	35	1.31	.74	.96	.15	.61	-.25	.28	.17	94.3	94.3	S5
9	2	35	1.31	.74	.99	.19	.82	.06	.20	.17	94.3	94.3	S9
2	3	35	.85	.62	1.00	.16	.81	-.09	.23	.20	91.4	91.4	S2
4	3	35	.85	.62	1.15	.46	1.47	.84	-.06	.20	91.4	91.4	S4
10	4	35	.52	.55	.95	-.01	.65	-.55	.36	.23	88.6	88.5	S10
1	6	35	.01	.47	1.22	.82	1.75	1.64	-.11	.27	82.9	82.8	S1
11	6	35	.01	.47	1.03	.21	.88	-.19	.27	.27	82.9	82.8	S11
15	8	35	-.38	.42	.98	-.03	.97	-.01	.32	.29	80.0	77.4	S15
13	10	35	-.71	.39	.93	-.34	1.04	.25	.37	.31	80.0	73.2	S13
8	11	35	-.86	.38	1.04	.27	1.13	.65	.25	.32	71.4	71.4	S8
12	11	35	-.86	.38	1.03	.26	1.00	.08	.29	.32	71.4	71.4	S12
6	12	35	-1.01	.38	.87	-.89	.80	-1.02	.51	.33	71.4	69.7	S6
3	13	35	-1.15	.37	.94	-.43	.89	-.61	.42	.33	68.6	68.0	S3
14	19	35	-1.94	.36	1.02	.22	.99	.00	.32	.34	62.9	64.3	S14
MEAN	7.4	35.0	.00	.53	1.00	.1	.95	.0			81.9	81.2	
P.SD	5.0	.0	1.08	.18	.09	.4	.32	.6			10.5	10.7	

Table 5. Results of validity analysis using the Rasch Model

C	Infit		Information	Outfit		Information
	MNSQ	ZSTD		MNSQ	ZSTD	
Item 1	0,95	0,25	Valid	0,40	- 0,20	Invalid
Item 2	0,96	0,15	Valid	0,61	-0,25	Valid
Item 3	0,99	0,19	Valid	0,82	0,06	Valid
Item 4	1,00	0,16	Valid	0,81	-0,09	Valid
Item 5	1,15	0,46	Valid	1,47	0,84	Valid
Item 6	0,95	-0,01	Valid	0,65	-0,55	Valid
Item 7	1,22	0,82	Valid	1,75	1,64	Invalid
Item 8	1,03	0,21	Valid	0,88	-0,19	Valid
Item 9	0,98	-0,03	Valid	0,97	-0,01	Valid
Item 10	0,93	-0,3	Valid	1,04	0,25	Valid
Item 11	1,04	0,27	Valid	1.13	0,65	Valid
Item 12	1,03	0,26	Valid	1,00	0,08	Valid
Item 13	0,87	-0,89	Valid	0,80	-1.02	Valid
Item 14	0,94	-0,43	Valid	0,89	-0,61	Valid
Item 15	1,02	0,22	Valid	0,99	0,00	Valid

Table 5 presents the results of the instrument validity test, enabling the assessment of whether any items are inappropriate. The findings indicate that all items are generally appropriate for evaluating students' conceptual understanding. Table 1 shows that, with the exception of items 1 and 7, which need revision or replacement due to their inappropriateness. The remaining items meet the necessary criteria. The Rasch analysis model aids instrument designers in making required adjustments to unsuitable items to ensure measurement reliability. Consequently, it can be concluded that the evaluation tools used in this research possess sufficient validity to measure students' conceptual understanding of the subject matter.

Table 6. Summary person and item reliability

SUMMARY OF 35 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	3.2	15.0	-1.74	.74	1.00	.08	.95	.11
SEM	.3	.0	.13	.02	.04	.13	.08	.13
P.SD	1.6	.0	.75	.14	.23	.79	.48	.76
S.SD	1.6	.0	.76	.14	.24	.80	.49	.77
MAX.	7.0	15.0	-.18	1.07	1.63	1.93	2.19	1.78
MIN.	1.0	15.0	-3.09	.58	.50	-2.49	.27	-2.08

	RMSE	TRUE SD	SEPARATION	Person RELIABILITY
REAL	.78	.00	.00	.00
MODEL	.75	.06	.08	.01

S.E. OF Person MEAN = .13

Person RAW SCORE-TO-MEASURE CORRELATION = .98
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .10 SEM = 1.48
STANDARDIZED (50 ITEM) RELIABILITY = .02

SUMMARY OF 15 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	7.4	35.0	.00	.53	1.00	.09	.95	.04
SEM	1.3	.0	.29	.05	.02	.10	.09	.17
P.SD	5.0	.0	1.08	.18	.09	.39	.32	.62
S.SD	5.2	.0	1.12	.19	.09	.40	.33	.64
MAX.	19.0	35.0	2.05	1.02	1.22	.82	1.75	1.64
MIN.	1.0	35.0	-1.94	.36	.87	-.89	.40	-1.02

	RMSE	TRUE SD	SEPARATION	Item RELIABILITY
REAL	.57	.92	1.63	.73
MODEL	.56	.93	1.66	.73

S.E. OF Item MEAN = .29

In this study, reliability is also evaluated to ensure that the test instrument yields consistent results when measuring students' conceptual understanding across different instances. Based on data analysis using Winsteps software, key values related to person reliability and item reliability are presented in Table 6. Table 6 shows the results of the Rasch Model analysis for the instrument measuring understanding of the concepts of motion and force. According to the reliability test analysis using the Rasch model, the results include Person Reliability and Item Reliability. The Item Reliability for this test is 0.73, which indicates high reliability according to the standards outlined in Table 3. This suggests that the item consistently measures what it is intended to measure and can be relied upon to assess the desired skill or knowledge.

However, when evaluating the overall reliability of all respondents, the score was only 0.01. This low value indicates that the majority of respondents experienced difficulties in answering the Force and Motion questions. The low person reliability score suggests that the respondents, as a group, found the questions challenging. This does not necessarily mean that the questions themselves are unreliable but rather that the students struggled with the material covered, in this case, Force and Motion. Low person reliability could be attributed to various factors, such as a lack of understanding of the material, ineffective teaching methods, or external factors like test anxiety. Consequently, improvements in instruction and learning strategies may be required to help students better understand and master the material, thereby enhancing the overall person reliability score. Individual reliability also showed a logit value of 0.01, while item reliability showed a value of 0.73, indicating that this instrument possesses high overall reliability.

Overall, the results of the analysis show that the instrument for understanding the concepts of motion and force exhibits good reliability and model fit. The instrument effectively differentiates between respondents' ability levels and item difficulty, as indicated by satisfactory reliability and separation values. This suggests that the questions are effective in assessing respondents' understanding of motion and force concepts. The measurement reliability for individuals was 0.80, and for items was 0.73, reflecting good consistency. Additionally, the good model fit was indicated by the Infit and Outfit MNSQ values being close to 1. This analysis provides evidence that the test instruments used can reliably measure the understanding of motion and force concepts in the tested population.

Based on Table 3, the item separation value reaches 1.66, indicating the assessment instrument's ability to distinguish between students who can answer the items correctly and those who cannot. To determine the grouping more thoroughly, the separation value is calculated using the Strata (*H*) equation, resulting in a value of 2.55, which can be rounded to 3. The analysis of respondents shows that the separation value obtained is only 0.08, with the H value reaching 0.44. The low *H* value of the respondents indicates that there are groups of students who still have difficulty understanding the material on force and motion. This highlights the importance of identifying and providing additional support to students who need it, as well as improving teaching strategies to ensure that the material is well understood by all students.

In Rasch model, item difficulty is typically assessed by considering parameters such as the probability of respondents with varying ability levels answering the item correctly. Therefore, the 'measure' section of the Rasch analysis provides insights into how difficult or easy an item is relative to the ability levels of the test respondents. Table 7 presents an analysis of item difficulty using the Rasch Model method. Based on Table 7, the value of each item can be seen from highest to lowest. Item number 7 has the highest value of 2.05, indicating that it is the most difficult question. Conversely, item number 14 has the lowest value of -1.94, indicating that it is the easiest question.

Table 7. Analysis of item difficulty using the Rasch model

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.
7	1	35	2.05	1.02
5	2	35	1.31	.74
9	2	35	1.31	.74
2	3	35	.85	.62
4	3	35	.85	.62
10	4	35	.52	.55
1	6	35	.01	.47
11	6	35	.01	.47
15	8	35	-.38	.42
13	10	35	-.71	.39
8	11	35	-.86	.38
12	11	35	-.86	.38
6	12	35	-1.01	.38
3	13	35	-1.15	.37
14	19	35	-1.94	.36
MEAN	7.4	35.0	.00	.53
P.SD	5.0	.0	1.08	.18

Based on the criteria given in Table 3, the items were divided into five categories according to their logit values. Items with a logit value higher than the standard deviation of logit (SD logit) are classified as "Very Difficult", which includes outliers. Examples are items numbers 7, 5, and 9. Meanwhile, items with a logit value between SD logit and half of SD logit are categorised as "Difficult". This indicates that these items require higher ability to answer. Examples include items numbers 4 and 2. Items with a logit value between -half the SD logit and half the SD logit are classified as "Moderate". Items in this category are considered to have a moderate level of difficulty. Item numbers 15, 11, 1, and 10 fall into this category. Items with a logit value between -half SD logit and -SD logit are categorised as "Easy", indicating that these items are relatively easier for respondents. Examples include items numbers 6, 12, 8, and 13. Items are categorised as outliers (Very Easy) when the logit value is more than -SD logit. Items numbers 3 and 14 fall into this category. The analysis revealed that the test items covered a range of difficulty levels, from the easiest to the most challenging, which is crucial for a thorough assessment. Including items that cater to different ability levels ensures that the test can accurately measure student understanding across the spectrum, from those with lower abilities to those with higher abilities (Linacre, 2016).

From Table 8, the error response shown by the subject can be observed. The results indicate that students with code 13 have low abilities but are able to answer questions with code 5, which is classified as very difficult. However, they answered questions with code 14, which are in the easy category, incorrectly. Therefore, it can be identified that student code 13 guessed the answer to question code 5, leading to the conclusion that they correctly answered question code 5 purely by guessing. Additionally, the figure illustrates a comparison between students with codes 14 and 29. If we only consider the number of correct answers, both students have the same ability. However, since student code 29 correctly answered more difficult questions, it can be concluded that student code 29 is smarter than student code 14. Hence, a person's ability is not only determined by the raw score but also by considering the difficulty level of the items, whether they are difficult, moderate, or easy to solve (Rasmuin & Luddin, 2022).

Table 8. Scalogram analysis from Guttman Scalogram

```
GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       |1   111 11
       |436823511024597
       |---------------
    33 +111111100000000  33
     3 +100101011000001  03
    14 +100011100010100  14
    29 +111110000100000  29
    32 +111001100000010  32
    26 +011100001100000  26
     6 +101110000000000  06
    13 +000001001010100  13
    27 +011000100100000  27
    28 +011001000100000  28
    34 +111010000000000  34
     2 +000011001000000  02
     7 +100100000001000  07
     8 +100100000010000  08
    15 +100010100000000  15
    19 +100100000000010  19
    20 +100000011000000  20
    22 +000000110001000  22
    23 +110000001000000  23
    24 +011001000000000  24
    30 +011010000000000  30
    31 +111000000000000  31
     4 +100000010000000  04
     5 +000110000000000  05
     9 +000000100001000  09
    12 +000101000000000  12
    16 +010000100000000  16
    17 +100010000000000  17
    21 +100010000000000  21
    25 +011000000000000  25
    35 +100000010000000  35
     1 +100000000000000  01
    10 +000000010000000  10
    11 +000100000000000  11
    18 +000001000000000  18
       |---------------
       |1   111 11
       |436823511024597
```

The results of this investigation suggest that the Rasch Model analysis can effectively account for both the quality of test takers and the items (Maulana et al., 2023). According to Isnani et al. (2019) and Susongko (2016), the Rasch Model provides insights into test taker consistency, including the likelihood of careless, guessing, or dishonest responses. This is undoubtedly highly beneficial for teachers in providing assessments that accurately reflect test takers' abilities (Mursidi & Soeharto, 2017; Suranata et al., 2021). The Rasch measurement model offers more precise information regarding the level of the items and individuals being assessed, as well as their suitability for the intended purposes (Tornabene et al., 2018). Furthermore, it aids teachers in examining, evaluating, and determining appropriate feedback for each student (Börkan, 2017), thus facilitating a more accurate and efficient assessment of students (Eerman Aslanoglu et al., 2020).

CONCLUSION

In conclusion, the FMCE used in this study demonstrated adequate overall validity and high reliability, with a logit value of 0.01 for individual reliability and 0.73 for item reliability. Analysis using Winstep Version 3.65.0 software identified certain items as outliers, while others were effectively categorised into difficulty levels. The instrument successfully differentiates between students of varying abilities, confirming its effectiveness in assessing understanding of the concept. However, the study's limitations, notably the suboptimal sample size, should be considered, as they may impact the robustness and generalisability of the findings. Despite these limitations, the quantitative approach offers valuable insights, though results should be interpreted with caution.

ACKNOWLEDGEMENTS

Thank you to all the MAN 2 Kuantan Singingi students who participated in this study, having studied the material on motion and force. Your involvement and cooperation have been invaluable in our efforts to develop and assess instruments for understanding these concepts. We hope that the results of this study will make a positive contribution to the field of education, particularly in physics learning.

REFERENCES

Auna Hidayati, R., & Taqwa, M. R. A. (2023). Computer based recitation program development with feed back in Newton’s Law topic. *International Journal of Education and Teaching Zone*, *2*(1), 113–129. https://doi.org/10.57092/ijetz.v2i1.49

Austvoll-Dahlgren, A., Semakula, D., Nsangi, A., Oxman, A. D., Chalmers, I., Rosenbaum, S., & Guttersrud, Ø. (2017). Measuring ability to assess claims about treatment effects: the development of the ‘Claim Evaluation Tools.’ *BMJ Open*, *7*(5), e013184. https://doi.org/10.1136/bmjopen-2016-013184

Bond, T. G., & Fox, C. M. (2019). Applying the Rasch model: Fundamental measurement in the human sciences (4th ed.). Routledge.

Börkan, B. (2017). Exploring variability sources in student evaluation of teaching via many-facet Rasch Model. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 15–15. https://doi.org/10.21031/epod.298462

Black, P., & Wiliam, D. (2020). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 32*(3), 171-195. https://doi.org/10.1007/s11092-020-09367-x

Browne, M. W., & Schreiber, J. B. (2020). The impact of conceptual understanding on problem-solving in physics education: A review. *International Journal of Science Education, 42*(9), 1427-1445. https://doi:10.1080/09500693.2020.1754557

Creswell, J. W. (2018). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research (4th Ed.)*. University of Nebraska–Lincoln.

Czerniak, C. M., & Johnson, J. A. (2020). Addressing misconceptions in physics: A focus on motion and force. *Journal of Research in Science Teaching, 57*(9), 1425-1443. https://doi:10.1002/tea.21624

Didik, L. A., Wahyudi, Muh., & Kafrawi, M. (2020). Identifikasi miskonsepsi dan tingkat pemahaman mahasiswa tadris fisika pada materi listrik dinamis menggunakan 3-tier diagnostic test. *Journal of Natural Science and Integration, 3*(2), 128. https://doi.org/10.24014/jnsi.v3i2.9911

Ding, L., & Beichner, R. J. (2021). Student understanding of force and motion: The role of integrating knowledge. *Journal of Research in Science Teaching, 58*(7), 1047-1069. https://doi:10.1002/tea.21612

Eerman A., Karakaya, I., & Sata, M. (2020). Evaluation of university students' rating behaviors in self and peer rating process via many facet Rasch Model. Eurasian *Journal of Educational Research, 20*(89), 1–22. https://doi.org/10.14689/ejer.2020.89.2

Finkelstein, N. D., & Leong, C. (2021). Conceptual difficulties and misconceptions in high school physics: Evidence from assessment and instruction. *Physics Education Research Conference Proceedings*, 2021, 207-212. https://doi:10.1119/perc.2021.pr.027

Hake, R. R., & Gibbons, A. S. (2020). Interactive engagement vs. traditional methods: How physics education research informs practice. *American Journal of Physics, 88*(1), 30-38. https://doi:10.1119/1.5123476

Hattie, J., & Timperley, H. (2020). The power of feedback for learning: Implications for assessment. *Review of Educational Research, 90*(4), 567-594. https://doi.org/10.3102/0034654319872322

Ismunandar, D., Rosyadi, R., Nurafifah, L., & Jofre, A. (2022). Improving students' problem solving abilities through the application of auditory intellectually repetition model. *Journal of Instructional Mathematics, 3*(2), 53–60. https://doi.org/10.37640/jim.v3i2.1385

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. https://doi.org/10.21831/reid.v5i2.20924

Kali, Y., & Ronen, M. (2019). Developing and evaluating assessments for conceptual understanding in science education. *Research in Science Education, 49*(3), 991-1010. https://doi.org/10.1007/s11165-018-9733-2

Kadir Masalesi, A. (2022). Resource identification and level of understanding of particle dynamics concepts. *International Journal of Education and Teaching Zone*, *1*(2), 204–221. https://doi.org/10.57092/ijetz.v1i2.48

McDermott, L. C., & Redish, E. F. (2021). Resource letter: PER-1: Physics education research. *American Journal of Physics*, *89*(6), 589-600. https://doi:10.1119/10.0002981

Mestre, J. P., Docktor, J. L., Strand, N. E., & Ross, B. H. (2011). Conceptual problem solving in physics (pp. 269–298). https://doi.org/10.1016/B978-0-12-387691-1.00009-0

Meyer, J. P., & Timm, W. R. (2021). Rasch analysis in education: Theory and practice. Educational Measurement: *Issues and Practice*, *40*(1), 15-29. https://doi.org/10.1111/emip.12324

Miller, K. L., & Kopp, M. D. (2020). Evaluating conceptual understanding in STEM education: The role of formative assessment. *Journal of Educational Measurement*, *57*(2), 159-177. https://doi.org/10.1111/jedm.12236

Misbah, M., Trisnowati, E., Rahim, A., & Zb, A. (2022). Investigating problem solving and mathematical connections in solving the fermi-dirac equation. *International Journal of Education and Teaching Zone*, *1*(2), 106–115. https://doi.org/10.57092/ijetz.v1i2.36

Mursidi, A., & Soeharto, S. (2017). An introduction: Evaluation of quality assurance for higher educational institutions using rasch model. *JETL (Journal of Education, Teaching and Learning)*, *1*(1), 1. https://doi.org/10.26737/jetl.v1i1.25

Nehru, N., Ananda, R., Zb, A., & Navalyyan, D. (2022). Analysis of mathematical critical thinking ability and mathematical creativity: Judging from the process of deriving the Fermi-Dirac Formula. *International Journal of Education and Teaching Zone*, 1(2), 87–96. https://doi.org/10.57092/ijetz.v1i2.33

Palimbong, J., Mujasam, M., & Allo, A. Y. T. (2019). Item analysis using Rasch model in semester final exam evaluation study subject in physics Class X TKJ SMK Negeri 2 Manokwari. *Kasuari: Physics Education Journal (KPEJ)*, *1*(1), 43–51. https://doi.org/10.37891/kpej.v1i1.40

Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, *15*(2), 20111. https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

Rasmuin, R., & Luddin, S. (2022). Tingkat kesulitan soal buatan guru bidang studi matematika menurut teori tes klasik pada tingkat SMP di Kota Baubau. *Jurnal Akademik Pendidikan Matematika*, 33–40. https://doi.org/10.55340/japm.v8i1.699

Ryan, Q. X., Frodermann, E., Heller, K., Hsu, L., & Mason, A. (2016). Computer problem-solving coaches for introductory physics: Design and usability studies. *Physical Review Physics Education Research*, *12*(1), 010105. https://doi.org/10.1103/PhysRevPhysEducRes.12.010105

Singh, C., & Chabay, R. (2022). Understanding and overcoming misconceptions in physics: A focus on motion and force. *Physical Review Physics Education Research, 18*(2), 020102. https://doi:10.1103/PhysRevPhysEducRes.18.020102

Sulman, F., Sutopo, S., & Kusairi, S. (2021). FMCE-PHQ-9 assessment with Rasch model in detecting concept understanding, cheating, and depression amid the covid-19 pandemic. *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah, 6*(2), 297–309. https://doi.org/10.24042/tadris.v6i2.9273

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan Rasch pada assessment pendidikan (1st ed.)*. Trim Komunikata.

Suranata, K., Dharsana, I. K., Paramartha, W. E., Dwiarwati, K. A., & Ifdil, I. (2021). Pengaruh model strength-based counseling dalam LMS Schoology untuk meningkatkan kesehatan mental dan kesejahteraan psikologis siswa. *JPPI (Jurnal Penelitian Pendidikan Indonesia), 7*(1), 137–146. https://doi.org/10.29210/020211142

Susongko, P. (2016). Validation of science achievement test with the Rasch model. *Jurnal Pendidikan IPA Indonesia*, *5*(2), 268–277. https://doi.org/10.15294/jpii.v5i2.7690

Tornabene, R. E., Lavington, E., & Nehm, R. H. (2018). Testing validity inferences for Genetic Drift Inventory scores using Rasch modeling and item order analyses. Evolution: *Education and Outreach, 11*(1), 6. https://doi.org/10.1186/s12052-018-0082-x

Wang, J., & Wang, Y. (2021). Improving conceptual understanding through diagnostic assessment and feedback. *Assessment & Evaluation in Higher Education, 46*(4), 533-547. https://doi.org/10.1080/02602938.2020.1798754

Yusuf, I., Zb, A., & Rozal, E. (2022). Understanding mathematical communication concepts and skills: analysis of the ability of prospective physics teachers? *International Journal of Education and Teaching Zone, 1*(2), 97–105. https://doi.org/10.57092/ijetz.v1i2.34

Zhu, Y., & Dong, Y. (2019). Misconceptions in elementary and high school students' understanding of motion and force: A review. *International Journal of Science Education, 41*(6), 732-750.