DOI: 10.14421/impulse.2025.51-06

IMPULSE: Journal of Research and Innovation in Physics Education
Volume 5, Issue 1, 69 – 83

© ISSN (p): 2798–1762; ISSN (e): 2798-1754
http://ejournal.uin-suka.ac.id/tarbiyah/impulse

Analyze the Students' Conceptual Understanding of Elasticity and Hooke's Law Using the Rasch Model

Jihan Ariqatur Rafiah[1*], Iqlides Ahmad Miyaqi[2]

[1] Master of Education, Indonesia International Islamic University, Indonesia
[2] Physics Education, Universitas Negeri Yogyakarta, Indonesia
*Corresponding author: jihan.rafiah@uiii.ac.id

ABSTRACT

This study investigates students' conceptual understanding of Elasticity and Hooke's Law by employing the Rasch measurement model. Conceptual understanding in physics is essential for enabling learners to apply principles across contexts and develop critical thinking skills. The 20 items multiple-choice test was developed and validated to assess mastery of elasticity-related concepts. The approach in this study was quantitative that conducted in Yogyakarta, Indonesia, involving 56 senior high students who had completed lessons on the relevant topic. The Rasch model was applied using the Ministeps software to analyze item difficulty and student ability. The Wright map and fit statistics indicated that most students had low to moderate levels.

INTISARI

Penelitian ini menyelidiki pemahaman konseptual siswa tentang Elastisitas dan Hukum Hooke dengan menggunakan model pengukuran Rasch. Pemahaman konseptual dalam fisika sangat penting untuk memungkinkan siswa menerapkan prinsip-prinsip dalam berbagai konteks dan mengembangkan keterampilan berpikir kritis. Tes pilihan ganda sebanyak 20 butir soal dikembangkan dan divalidasi untuk menilai penguasaan konsep-konsep yang berhubungan dengan elastisitas. Pendekatan dalam penelitian ini adalah kuantitatif yang dilakukan di Yogyakarta, Indonesia, dengan melibatkan 56 siswa sekolah menengah atas yang telah menyelesaikan pelajaran tentang topik yang relevan. Model Rasch diterapkan dengan menggunakan perangkat lunak Ministeps untuk menganalisis tingkat kesulitan soal dan kemampuan siswa. Peta Wright dan statistik kecocokan menunjukkan bahwa sebagian besar siswa memiliki tingkat kemampuan rendah hingga sedang.

ARTICLE HISTORY
Received: June 9, 2025
Accepted: September 25, 2025

KEYWORDS:

Conceptual Understanding, Elasticity and Hooke's Law, Multiple Choice Questions (MCQs), Rasch Model

KATA KUNCI:

Elastisitas dan Hukum Hook, Pemahaman Konsep, *Rasch Model*, Soal Pilihan Ganda

A. Introduction

Understanding physical concepts is crucial for forming a strong knowledge base and imparting problem-solving and critical thinking skills to students. As physics is a natural science discipline, knowing basic concepts such as force, motion, energy, and physical laws is important. The understanding of these concepts in the sense of

* Corresponding author:
Jihan Ariqatur Rafiah, Master of Education, Indonesia International Islamic University, Indonesia
✉ jihan.rafiah@uiii.ac.id

definitions and operational usage, as well as perceptions of real-world observations, is needed. The study by Phanphech [1] stated that physics teaching needs to include explanation of concepts and use of physics concepts in situations where the students can relate them to enhance deeper understanding. Concept understanding can also be employed to refer to practical understanding when accompanied by rational ability to apply knowledge in the setting where they learned it [2]. Furthermore, understanding physics concepts, such as elasticity and Hooke's Law, is essential for students, as these form the foundation for other learning in mechanics and engineering. Despite the importance of these issues, they are challenging for students, with the resulting long-term misconceptions precluding deeper understanding. Traditional testing places higher emphasis on memorization than on the knowledge of the concept, thereby unable to identify some areas where students are perplexed.

Evaluating students' conceptual knowledge is an important part of determining their level of comprehension of kinematics material. It can also help teachers prepare more effective learning strategies, identify struggling students, and improve the quality of physics education. The Rasch model is one of the measurement methods that can be used in assessing students' comprehension of Elasticity and Hooke's Law. The Rasch model has advantages like high objectivity, rich detailed information about the mastery level of each concept, and ease of data analysis and interpretation [3]. The Rasch model also forecasts the probability of correct answers to a test based on ratings of two facets, item difficulty and individual ability, along the continuum between the two [4]. Rasch model analysis improves the accuracy and quality of tests and studies because it also allows the creation of a variety of different measurement tools. The Rasch model is used as a reference in analysing the response structure, rather than simply to describe the statistics of a given response. The ultimate goal is to achieve the maximum level of accuracy and objectivity in the measurement process, thus allowing for a more precise relationship between the measurement instrument and the essential characteristics of the individual. In the context of the Rasch model, this 'settled' scoring pattern is nothing but a measurement whose results depend on who is being measured (test-dependent scoring); whereas what must be done in quantitative research in the social sciences is objective measurement [5]. Analysis using the Rasch model also evaluates a measurement instrument's validity and reliability [6].

Understanding key physics concepts is fundamental to the development of scientific literacy and critical thinking skills among students. Physics, as a core discipline in natural science, requires not only memorization of formulas but also a deep conceptual grasp of ideas such as force, motion, energy, and physical laws. Among these, elasticity and Hooke's Law are central to understanding how materials behave under stress, making them foundational for further studies in mechanics and engineering. Conceptual understanding, as opposed to procedural or rote learning, enables learners to apply their knowledge in varied contexts, fostering transfer of learning and long-term retention [7]. However, research has consistently shown that students struggle with abstract physics topics, including elasticity. This challenge often leads to misconceptions that persist despite instruction. Conceptual understanding involves more than recalling definitions; it includes the ability to use concepts to explain, analyze, and make sense of real-world phenomena [1]. This practical dimension of understanding being able to apply knowledge rationally in context is what distinguishes meaningful learning from surface-level acquisition [2]. Elasticity, defined as a material's ability to return to its original shape after deformation, and Hooke's Law, which describes the linear relationship between force and displacement ($F = kx$), are not only important physics principles but also tools to explain everyday phenomena and technological applications.

Despite their importance, traditional assessment methods are often insufficient to evaluate conceptual understanding. Standard tests tend to emphasize factual recall and algorithmic problem solving, which may obscure deeper learning gaps. As Hestenes, Wells, and Swackhamer [8] highlighted, students who can explain physical phenomena using conceptual reasoning outperform those who rely on formulaic procedures. Instructional approaches such as inquiry-based labs, peer instruction, and the use of simulations have been shown to improve conceptual grasp, particularly in abstract topics like elasticity [9]. But evaluating the impact of such teaching strategies requires robust tools that can accurately measure student understanding.

One such tool is the multiple-choice question (MCQ). MCQs are widely used in education due to their efficiency, scalability, and objectivity. Properly constructed MCQs, with plausible distractors based on known misconceptions, can assess not only recall but also higher-order cognitive skills [10]. Their standardized format makes them ideal for large-scale assessment, and when combined with advanced psychometric models such as the Rasch model, they become powerful diagnostic instruments [11].

The Rasch model is a probabilistic model under item response theory (IRT) that converts raw ordinal scores into interval-level measures, allowing for precise analysis of both student ability and item [3]. Unlike classical test theory, the Rasch model places students and items on the same linear continuum, providing deeper insight into how well items discriminate between levels of understanding. It allows for the development of fair and valid tests by identifying misfitting items and unusual response patterns. As Boone, Staver, and Yale [12] explain, Rasch analysis helps to refine assessments by ensuring that items align with the intended construct and function consistently across groups.

In the context of science education, particularly physics, Rasch modeling is highly relevant. Concepts like elasticity and Hooke's Law involve abstract reasoning and proportional thinking that many students find difficult. The Rasch model allows educators to determine which items reveal conceptual understanding and which distractors may be misleading or ineffective. It also enables the identification of students whose responses deviate from expected patterns, possibly due to guessing, misunderstanding, or disengagement. Moreover, Rasch modeling emphasizes measurement invariance ensuring that item parameters remain stable regardless of the student sample and person independence ensuring student ability measures are not dependent on the specific items used, as long as the model fits [5], [6].

These advantages make the Rasch model especially useful in developing conceptual diagnostic tools. For instance, research by Zacharia and Anderson [9] demonstrated that visualizing the force-displacement relation in elasticity through simulation and experimentation helped students build more robust conceptual frameworks. Such insights can be integrated into test design by using the Rasch model to calibrate item difficulty and detect learning gaps that traditional tests may miss.

Therefore, this study aims to evaluate students' conceptual understanding of elasticity and Hooke's Law through a multiple-choice diagnostic test, applying the Rasch measurement model to analyze both item quality and student ability more effectively than traditional assessments. By identifying patterns in student responses and aligning item difficulty with conceptual demands, this research seeks to contribute to the development of more valid, reliable, and informative assessments in physics education.

B. Method

This study utilises a quantitative descriptive design to explore students' conceptual understanding of Hooke's Law and Elasticity as measured by the Rasch measurement model. The Rasch model is an item response theory (IRT) one-parameter logistic model that is useful for calibrating ordinal raw scores to interval-level data, allowing both item difficulty and student ability to be analyzed on the same linear scale [12]. The subjects of this study were 56 students from grades XI and XII at MA Ali Maksum Yogyakarta, who had completed lessons on Elasticity and Hooke's Law. The sample was selected using purposive sampling to ensure that participants had studied the relevant material. The participants included 32 female and 24 male students, with an age range of 16 to 18 years old. All students were enrolled in the science stream and had participated in physics classes covering the target concepts. This demographic information helps to contextualize the findings, particularly in relation to patterns of understanding that may differ across age or gender groups.

The instrument was a 20-item conceptual test consisting of multiple-choice items with five response options, including one correct answer and four plausible distractors. The items were developed based on key concepts of the physics syllabus, such as Young's modulus, elastic and plastic deformation, Hooke's Law, and force-extension graphs [13]. Each distractor was designed to reflect common student misconceptions, allowing the test to effectively diagnose conceptual understanding rather than rote memorization or procedural recall. Prior to administration, the instrument underwent content validation by physics education experts and statistical validation of reliability and item functioning through Rasch model analysis.

For data analysis, the Rasch model was utilized with the Ministep program. Item and person reliability and separation indices were evaluated to assess the test's precision and ability to differentiate between different student knowledge levels. Student ability and item difficulty were in terms of logits. Fit statistics (infit and outfit mean square values) were used for model-data fit assessment, for which values ranging from 0.5 to 1.5 are acceptable [13]. A Wright map (person-item map) was made for a visual display of the distribution of the items and the abilities of the students on a common scale.

C. Results and Discussion

In line with the aim of this study to evaluate students' conceptual understanding of elasticity and Hooke's Law using a Rasch-based diagnostic test, this section presents the results of the analysis by focusing on how students' abilities align with the conceptual demands of the test items. The Rasch model allows for both item difficulty and student ability to be placed on a single linear scale (logit), making it possible to group students according to their level of conceptual understanding. The following table categorizes students into four levels of conceptual understanding based on their Rasch ability estimates:

Table 1. Categorization of Students' Conceptual Understanding Based on Rasch Ability Scores

Logit Range	Number of Students	Level of Understanding	Description
> +1.0	1	High	Demonstrates accurate and transferable understanding
0.0 to +1.0	18	Moderate	Understands key concepts with occasional inconsistencies
-1.0 to 0.0	29	Low	Shows partial understanding with significant misconceptions
< -1.0	8	Very Low	Lacks conceptual clarity, responses indicate guessing
Total	56		

A method that has been used is the Rasch model analysis, in which item validity can be established through three overall measures: Outfit Mean Square (MNSQ), Z-standardized (ZSTD), and Point Measure Correlation (CORR). An MNSQ value is considered valid when it falls between 0.5 to 1.5, i.e., the item has a proper level of fit with the model. On the other hand, a ZSTD value is deemed appropriate if it falls between -2.0 and +2.0, as this would mean no significant statistical deviation. A

CORR value must be greater than 0.2 because it refers to positive correlation of the item with the measured ability. The three parameters serve as the basis to determine whether an item is appropriate to use as a research component or needs revision, or can be eliminated. Figure 1 presents the item validity analysis results like Outfit MNSQ, ZSTD, and Point Measure Correlation value. Figure 1 presents the item validity analysis results, like Outfit MNSQ, ZSTD, and Point Measure Correlation values.

ENTRY	TOTAL	TOTAL		MODEL	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH			
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Item	
16	5	56	2.25	.47	.99	.09	.96	.05	.18	.14	91.1	91.1	S16	VALID
8	10	56	1.43	.36	1.08	.42	1.11	.47	.04	.19	82.1	82.1	S8	VALID
17	10	56	1.43	.36	.91	-.37	1.01	.11	.33	.19	82.1	82.1	S17	VALID
18	13	56	1.09	.32	.92	-.41	1.00	.09	.32	.21	78.6	76.8	S18	VALID
3	15	56	.89	.31	1.09	.64	1.17	.95	.02	.22	75.0	73.4	S3	VALID
20	17	56	.71	.30	.90	-.80	.94	-.34	.39	.23	71.4	70.0	S20	VALID
5	19	56	.53	.29	1.03	.34	1.10	.83	.14	.23	69.6	66.9	S5	VALID
2	20	56	.45	.29	1.01	.19	1.03	.30	.20	.23	60.7	65.6	S2	VALID
4	20	56	.45	.29	1.07	.70	1.17	1.45	.07	.23	60.7	65.6	S4	VALID
6	21	56	.37	.28	1.06	.64	1.08	.73	.12	.24	66.1	64.2	S6	VALID
11	21	56	.37	.28	1.16	1.72	1.18	1.62	-.06	.24	58.9	64.2	S11	VALID
12	21	56	.37	.28	.98	-.18	.95	-.39	.28	.24	62.5	64.2	S12	VALID
19	28	56	-.17	.28	.93	-1.04	.93	-.95	.37	.25	66.1	59.6	S19	VALID
13	33	56	-.56	.28	.93	-.91	.91	-.95	.38	.24	71.4	62.5	S13	VALID
9	34	56	-.64	.28	.97	-.38	.99	-.09	.29	.24	73.2	63.6	S9	VALID
15	36	56	-.80	.29	.98	-.21	1.00	.02	.27	.24	67.9	66.0	S15	VALID
7	39	56	-1.06	.30	1.07	.63	1.18	1.22	.05	.23	67.9	70.5	S7	VALID
10	42	56	-1.34	.32	.90	-.59	.91	-.41	.39	.22	80.4	75.1	S10	VALID
14	51	56	-2.61	.47	.92	-.08	.78	-.39	.32	.15	91.1	91.1	S14	VALID
1	53	56	-3.17	.60	.88	-.06	.47	-.90	.45	.12	94.6	94.6	S1	OVERVIT
MEAN	25.4	56.0	.00	.33	.99	.0	.99	.2			73.6	72.5		
P.SD	13.2	.0	1.30	.08	.08	.6	.16	.8			10.4	10.3		

Figure 1. Item Validity

Based on Figure 1, it can be concluded that all items are valid and can be used in the study. However, further examination is needed for item number 1, as its Outfit MNSQ value is below 0.5, indicating a potential overfit. An overfit item refers to one that fits the model too perfectly, which may raise concerns regarding its quality. This condition could suggest that the item is too easy or that respondents answered it in a highly uniform manner, reducing its ability to differentiate between varying levels of respondent ability. Although the item still falls within acceptable limits in terms of ZSTD and CORR values, its unusually low Outfit MNSQ warrants closer scrutiny before making a final decision about its inclusion in the instrument.

The Wright map showed that most students' ability measures clustered between -1.0 and +1.0 logits, indicating low to moderate conceptual understanding. Only one student achieved a high ability level (logit > +1.0). In terms of item difficulty, several items were identified as highly challenging, yet valid and informative. One such example is item 16, which assesses students' understanding of how dimensions affect material elongation:

16. Seorang siswa Aliyah melakukan percobaan terhadap empat buah kawat terbuat dari bahan identik dengan data sebagai berikut:

Kawat	Panjang (cm)	Diameter (mm)
1	50	0,5
2	100	1
3	200	2
4	300	3

Manakah kawat yang memiliki pertambahan panjang terbesar apabila diberi gaya yang sama?

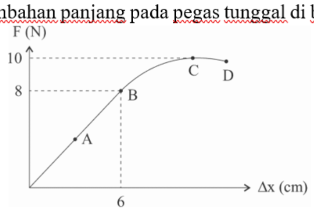
A. Kawat 1 karena rasio diameter terkecil
B. Kawat 2 karena rasio diameter yang setara
C. Kawat 3 karena rasio diameter sebesar seribu
D. Kawat 4 karena memiliki rasio diameter terbesar
E. Tidak ada karena tiap kawat berasio sama

Figure 2. Item Number 16

The correct answer was A, yet many students selected distractors B or D, reflecting common misconceptions about material deformation. Rasch analysis revealed that this item had a difficulty logit of +1.85, an Outfit MNSQ of 0.95, and a CORR of +0.35, confirming its status as a valid and informative item. It effectively distinguished students who grasped the conceptual relationship between length, cross-sectional area, and strain.

In addition, item S8 examined student reasoning about shifting the elastic limit on a spring's force-extension graph. It asked:

8. Perhatikan grafik hubungan gaya dengan pertambahan panjang pada pegas tunggal di bawah.



Apabila gaya diperbesar hingga melewati titik C, maka pegas akan bersifat plastis dan tidak kembali ke panjang semula. Langkah yang dapat dilakukan agar titik C dapat berpindah ke titik B adalah . . .

A. Memperbesar panjang pegas
B. Memperkecil konstanta pegas
C. Memotong pegas menjadi dua bagian
D. Memperbesar konstanta pegas
E. Mengganti pegas dengan susunan paralel

Figure 3. Item Number 8

(Correct answer: D. Memperbesar konstanta pegas)

This item had a logit of +1.65, was statistically fit (MNSQ = 1.02), and had CORR = +0.41. The distractors functioned well and revealed that many students lacked an integrated understanding of graphical and conceptual representations of elastic behavior. Another challenging item, S17, tested students' procedural reasoning in determining the force in a complex spring system. It had the highest item difficulty (+2.10 logits), a fit statistic of 0.94, and CORR = +0.39, validating its use for identifying higher-order thinkers.

These examples not only illustrate the effectiveness of Rasch analysis in evaluating test quality but also emphasize the value of carefully constructed multiple-choice items in diagnosing students' conceptual understanding. The findings support the conclusion that the Rasch-based diagnostic test is a powerful tool for revealing learning gaps. Therefore, strengthening both test development and instructional strategies is crucial for enhancing physics learning outcomes related to elasticity and Hooke's Law.

The Rasch model analysis revealed a significant difference between person and item reliability. Person reliability was very low (0.00), indicating that the instrument could not effectively distinguish between students with different ability levels. This suggests that most students gave similar responses, which may be due to the test being too easy or too difficult, or a lack of variation in student ability. In contrast, item reliability was very high (0.93), showing that the test items varied well in difficulty and could be distinguished effectively. However, a negative correlation between item scores and measures (-0.99) indicates a potential issue in scoring, such as a reversed scoring pattern. These findings highlight the need for improving the test design and reviewing the scoring process to ensure accurate and reliable assessment of students' conceptual understanding. The reliability results can be seen in Figure 4 and Figure 5.

```
SEPARATION   .00  Person RELIABILITY  .00
SEPARATION   .04  Person RELIABILITY  .00
```

Figure 4. Measured Person (Reliability)

```
SEPARATION  3.63  Item   RELIABILITY  .93
SEPARATION  3.67  Item   RELIABILITY  .93
```

Figure 5. Measured Item (Reliability)

Data analysis from the measurement of the understanding of the concept of elasticity and Hook's law was performed by applying the Rasch model using Winstep Rasch software. The level of students' concept understanding was analysed based on the output results, specifically through Person Measure and Person Fit Order. In addition, to illustrate the achievement of concept understanding, the Wright Map presented in Table 1 was used. On the Wright Map, the vertical discontinuous line illustrates the distribution of participants and items from the lowest to the highest level of ability or difficulty (arranged from bottom to top). Students at the bottom of the map indicate a low level of concept mastery, while those above the vertical line indicate a higher level of concept understanding. The symbol 'M' indicates the mean score, while "S" and 'T' represent one and two standard deviations from the mean, respectively. Figure 6 shows the Wright map:

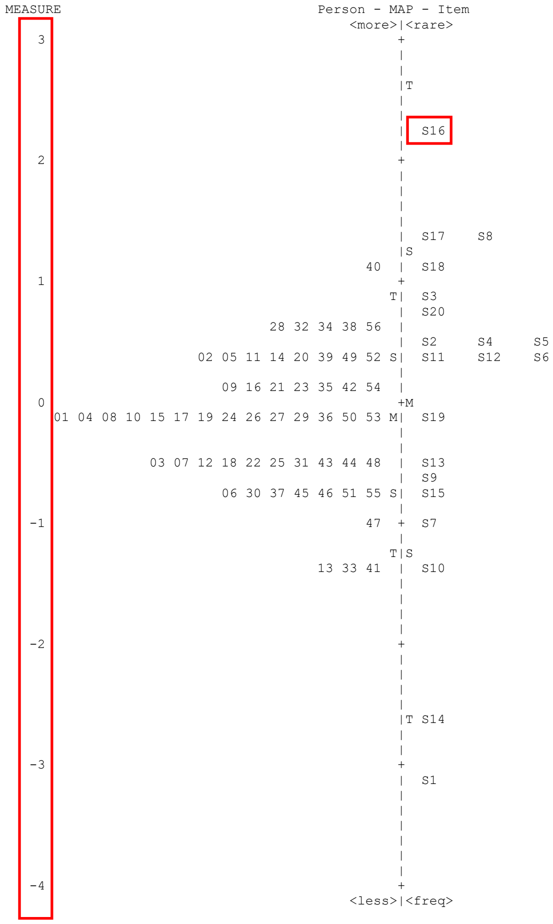


Figure 6. Wright Map

Based on the results of the Wright Map analysis in Figure 1, the ability of students to understand the concepts of the tested material is analyzed through the distribution of logit positions between students and the difficulty level of the question items. The logit scale is used to map the ability (person) on the left side and the difficulty level of the question (item) on the right side, with a midpoint of 0.0 as the average. The mapping results showed that most learners were in the logit range between -1 and +1, indicating that their abilities were still at a moderate level and tended to be below average. Only a few learners had high ability, such as student number 40, who was above logit +1. On the question side, item S16 was at the top, indicating that it was the most difficult question in the instrument. Meanwhile, the easiest item was S1, which is located at the lowest logit position, which is below -3. From this distribution, it can be seen that there are still many learners who are not able to answer items with high difficulty levels, such as S16, S8, and S17, because their logit position is far below these questions. Thus, this Wright Map shows that students' conceptual ability in understanding the material still needs to be improved, especially in dealing with questions of high complexity.

Based on the 2023 National Assessment results released by the Centre for Educational Assessment (Pusmendik) [14], several schools in the Special Region of Yogyakarta showed high literacy and numeracy achievements. This is reflected in the report 'Profile of Education Units with High AKM Achievement', which notes that a number of schools in Yogyakarta fall into the category of high AKM achievement in 2023. Compared to the findings in this study, where most students are in the low to moderate concept understanding category, there is a gap in achievement that needs attention. This shows that although Yogyakarta is known to have a good quality of education in general, not all educational units have shown optimal achievement results in physics concepts, especially in elasticity and Hooke's Law.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	JMLE MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
40	14	20	1.15	.54	1.09	.45	1.12	.39	.33	.40	75.0	73.9	40
28	12	20	.60	.51	.64	-2.09	.53	-1.08	.68	.44	90.0	69.5	28
32	12	20	.60	.51	1.21	1.10	1.16	.48	.30	.44	60.0	69.5	32
34	12	20	.60	.51	1.40	1.95	1.49	1.07	.16	.44	50.0	69.5	34
38	12	20	.60	.51	1.26	1.32	1.21	.58	.27	.44	50.0	69.5	38
56	12	20	.60	.51	1.16	.85	1.06	.28	.35	.44	60.0	69.5	56
2	11	20	.34	.51	1.04	.25	.91	-.09	.45	.46	65.0	68.9	02
5	11	20	.34	.51	.63	-2.20	.53	-1.25	.71	.46	85.0	68.9	05
11	11	20	.34	.51	.64	-2.12	.54	-1.22	.70	.46	85.0	68.9	11
14	11	20	.34	.51	1.27	1.37	1.33	.88	.27	.46	65.0	68.9	14
20	11	20	.34	.51	1.02	.18	.90	-.11	.46	.46	65.0	68.9	20
39	11	20	.34	.51	1.30	1.52	1.17	.53	.28	.46	55.0	68.9	39
49	11	20	.34	.51	1.52	2.43	1.52	1.23	.11	.46	45.0	68.9	49
52	11	20	.34	.51	1.14	.75	1.05	.25	.38	.46	65.0	68.9	52
9	10	20	.08	.51	.61	-2.23	.52	-1.43	.73	.47	90.0	70.6	09
16	10	20	.08	.51	.61	-2.23	.52	-1.43	.73	.47	90.0	70.6	16
21	10	20	.08	.51	.77	-1.20	.66	-.91	.63	.47	80.0	70.6	21
23	10	20	.08	.51	.69	-1.69	.59	-1.18	.68	.47	80.0	70.6	23
35	10	20	.08	.51	1.46	2.10	1.89	2.02	.12	.47	50.0	70.6	35
42	10	20	.08	.51	1.19	.97	1.06	.29	.37	.47	60.0	70.6	42
54	10	20	.08	.51	1.34	1.63	1.25	.73	.26	.47	50.0	70.6	54
1	9	20	-.18	.52	.56	-2.36	.48	-1.65	.77	.49	90.0	72.0	01
4	9	20	-.18	.52	.71	-1.47	.60	-1.18	.69	.49	80.0	72.0	04
8	9	20	-.18	.52	.65	-1.84	.55	-1.37	.72	.49	90.0	72.0	08
10	9	20	-.18	.52	.65	-1.84	.55	-1.37	.72	.49	90.0	72.0	10
15	9	20	-.18	.52	.87	-.59	.76	-.60	.58	.49	80.0	72.0	15
17	9	20	-.18	.52	.88	-.50	.84	-.35	.56	.49	80.0	72.0	17
19	9	20	-.18	.52	.78	-1.06	.66	-.96	.64	.49	80.0	72.0	19
24	9	20	-.18	.52	.56	-2.36	.48	-1.65	.77	.49	90.0	72.0	24
26	9	20	-.18	.52	.65	-1.84	.55	-1.37	.72	.49	90.0	72.0	26
27	9	20	-.18	.52	.56	-2.36	.48	-1.65	.77	.49	90.0	72.0	27
29	9	20	-.18	.52	.56	-2.36	.48	-1.65	.77	.49	90.0	72.0	29
36	9	20	-.18	.52	1.24	1.13	1.19	.60	.34	.49	60.0	72.0	36
50	9	20	-.18	.52	1.09	.48	1.30	.87	.40	.49	70.0	72.0	50
53	9	20	-.18	.52	1.27	1.25	1.19	.62	.32	.49	60.0	72.0	53
3	8	20	-.45	.53	.84	-.64	.95	-.02	.58	.50	85.0	73.5	03
7	8	20	-.45	.53	.86	-.54	.77	-.55	.59	.50	85.0	73.5	07
12	8	20	-.45	.53	1.10	.49	1.07	.31	.44	.50	65.0	73.5	12
18	8	20	-.45	.53	.84	-.64	.95	-.02	.58	.50	85.0	73.5	18
22	8	20	-.45	.53	.72	-1.22	.61	-1.09	.68	.50	85.0	73.5	22
25	8	20	-.45	.53	.67	-1.52	.55	-1.31	.72	.50	85.0	73.5	25
31	8	20	-.45	.53	.65	-1.65	.54	-1.38	.73	.50	85.0	73.5	31
43	8	20	-.45	.53	1.21	.92	1.20	.64	.36	.50	65.0	73.5	43
44	8	20	-.45	.53	1.13	.61	1.13	.46	.41	.50	65.0	73.5	44
48	8	20	-.45	.53	.96	-.10	.82	-.40	.54	.50	75.0	73.5	48
6	7	20	-.74	.54	.64	-1.47	.53	-1.28	.74	.50	90.0	75.3	06
30	7	20	-.74	.54	.73	-1.03	.59	-1.04	.69	.50	80.0	75.3	30
37	7	20	-.74	.54	1.50	1.73	1.57	1.34	.18	.50	60.0	75.3	37
45	7	20	-.74	.54	1.44	1.57	1.64	1.46	.20	.50	60.0	75.3	45
46	7	20	-.74	.54	1.40	1.43	1.59	1.37	.23	.50	70.0	75.3	46
51	7	20	-.74	.54	1.25	.97	1.26	.73	.34	.50	70.0	75.3	51
55	7	20	-.74	.54	1.33	1.22	1.94	1.96	.24	.50	70.0	75.3	55
47	6	20	-1.05	.57	1.20	.72	1.17	.49	.39	.51	70.0	77.9	47

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%	Person
13	5	20	-1.39	.60	.80	-.50	.67	-.49	.64	.51	85.0	81.3	13
33	5	20	-1.39	.60	2.00	2.37	2.28	1.89	-.13	.51	65.0	81.3	33
41	5	20	-1.39	.60	2.16	2.65	2.69	2.28	-.25	.51	65.0	81.3	41
MEAN	9.1	20.0	-.18	.53	1.01	-.1	.99	-.1			73.6	72.5	
P.SD	1.9	.0	.53	.02	.36	1.5	.48	1.1			13.3	3.0	

Figure 7. Person Measure

The Person Measure table in Figure 5 shows the logit value that describes the level of students' ability to understand the tested material. Based on the analysis results, the average logit value is -0.18 with a standard deviation of 0.53. According to Kurli et al [15], the standard deviation value can be used as a reference to classify students' abilities into high, medium, and low categories. Learners who have a logit value above +0.35 (mean + 1 SD) are classified in the high category, but out of 56 students, only one reached this category, student number 40, with a logit value of 1.15 and a total score of 14. This shows that the majority of participants still do not show a strong understanding of the material. Learners with logits above 0, such as students 28, 32, 34, 38, and 56, can be categorised as being at a medium level of ability. Meanwhile, most participants were below average, with negative logit values, such as students 01, 08, 11, 16, 19, 21, and 26. There were even participants with very low logit values, such as participants 13, 33, and 41, with a logit of -1.39, indicating very weak concept understanding. Based on this distribution, it can be concluded that the majority of learners in this study were in the low to medium ability category, and only a few reached the high category. This indicates the need for more effective learning interventions to improve learners' understanding of the tested material.

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%	Person
41	5	20	-1.39	.60	2.16	2.65	2.69	2.28	A-.25	.51	65.0	81.3	41
33	5	20	-1.39	.60	2.00	2.37	2.28	1.89	B-.13	.51	65.0	81.3	33
55	7	20	-.74	.54	1.33	1.22	1.94	1.96	C .24	.50	70.0	75.3	55
35	10	20	.08	.51	1.46	2.10	1.89	2.02	D .12	.47	50.0	70.6	35
45	7	20	-.74	.54	1.44	1.57	1.64	1.46	E .20	.50	60.0	75.3	45
46	7	20	-.74	.54	1.40	1.43	1.59	1.37	F .23	.50	70.0	75.3	46
37	7	20	-.74	.54	1.50	1.73	1.57	1.34	G .18	.50	60.0	75.3	37
49	11	20	.34	.51	1.52	2.43	1.52	1.23	H .11	.46	45.0	68.9	49
34	12	20	.60	.51	1.40	1.95	1.49	1.07	I .16	.44	50.0	69.5	34
54	10	20	.08	.51	1.34	1.63	1.25	.73	J .26	.47	50.0	70.6	54
14	11	20	.34	.51	1.27	1.37	1.33	.88	K .27	.46	65.0	68.9	14
39	11	20	.34	.51	1.30	1.52	1.17	.53	L .28	.46	55.0	68.9	39
50	9	20	-.18	.52	1.09	.48	1.30	.87	M .40	.49	70.0	72.0	50
53	9	20	-.18	.52	1.27	1.25	1.19	.62	N .32	.49	60.0	72.0	53
38	12	20	.60	.51	1.26	1.32	1.21	.58	O .27	.44	50.0	69.5	38
51	7	20	-.74	.54	1.25	.97	1.26	.73	P .34	.50	70.0	75.3	51
36	9	20	-.18	.52	1.24	1.13	1.19	.60	Q .34	.49	60.0	72.0	36
32	12	20	.60	.51	1.21	1.10	1.16	.48	R .30	.44	60.0	69.5	32
43	8	20	-.45	.53	1.21	.92	1.20	.64	S .36	.50	65.0	73.5	43
47	6	20	-1.05	.57	1.20	.72	1.17	.49	T .39	.51	70.0	77.9	47
42	10	20	.08	.51	1.19	.97	1.06	.29	U .37	.47	60.0	70.6	42
56	12	20	.60	.51	1.16	.85	1.06	.28	V .35	.44	60.0	69.5	56
52	11	20	.34	.51	1.14	.75	1.05	.25	W .38	.46	65.0	68.9	52
44	8	20	-.45	.53	1.13	.61	1.13	.46	X .41	.50	65.0	73.5	44
40	14	20	1.15	.54	1.09	.45	1.12	.39	Y .33	.40	75.0	73.9	40
12	8	20	-.45	.53	1.10	.49	1.07	.31	Z .44	.50	65.0	73.5	12
18	8	20	-.45	.53	.84	-.64	.95	-.02	z .58	.50	85.0	73.5	18
17	9	20	-.18	.52	.88	-.50	.84	-.35	y .56	.49	80.0	72.0	17
15	9	20	-.18	.52	.87	-.59	.76	-.60	x .58	.49	80.0	72.0	15
7	8	20	-.45	.53	.86	-.54	.77	-.55	w .59	.50	85.0	73.5	07
13	5	20	-1.39	.60	.80	-.50	.67	-.49	v .64	.51	85.0	81.3	13
19	9	20	-.18	.52	.78	-1.06	.66	-.96	u .64	.49	80.0	72.0	19
21	10	20	.08	.51	.77	-1.20	.66	-.91	t .63	.47	80.0	70.6	21
30	7	20	-.74	.54	.73	-1.03	.59	-1.04	s .69	.50	80.0	75.3	30
22	8	20	-.45	.53	.72	-1.22	.61	-1.09	r .68	.50	85.0	73.5	22

4	9	20	-.18	.52	.71	-1.47	.60	-1.18	q	.69	.49	80.0	72.0	04
23	10	20	.08	.51	.69	-1.69	.59	-1.18	p	.68	.47	80.0	70.6	23
25	8	20	-.45	.53	.67	-1.52	.55	-1.31	o	.72	.50	85.0	73.5	25
8	9	20	-.18	.52	.65	-1.84	.55	-1.37	n	.72	.49	90.0	72.0	08
10	9	20	-.18	.52	.65	-1.84	.55	-1.37	m	.72	.49	90.0	72.0	10
26	9	20	-.18	.52	.65	-1.84	.55	-1.37	l	.72	.49	90.0	72.0	26
31	8	20	-.45	.53	.65	-1.65	.54	-1.38	k	.73	.50	85.0	73.5	31
6	7	20	-.74	.54	.64	-1.47	.53	-1.28	j	.74	.50	90.0	75.3	06
11	11	20	.34	.51	.64	-2.12	.54	-1.22	i	.70	.46	85.0	68.9	11
28	12	20	.60	.51	.64	-2.09	.53	-1.08	h	.68	.44	90.0	69.5	28
5	11	20	.34	.51	.63	-2.20	.53	-1.25	g	.71	.46	85.0	68.9	05
9	10	20	.08	.51	.61	-2.23	.52	-1.43	f	.73	.47	90.0	70.6	09
16	10	20	.08	.51	.61	-2.23	.52	-1.43	e	.73	.47	90.0	70.6	16
1	9	20	-.18	.52	.56	-2.36	.48	-1.65	d	.77	.49	90.0	72.0	01
24	9	20	-.18	.52	.56	-2.36	.48	-1.65	c	.77	.49	90.0	72.0	24
27	9	20	-.18	.52	.56	-2.36	.48	-1.65	b	.77	.49	90.0	72.0	27
29	9	20	-.18	.52	.56	-2.36	.48	-1.65	a	.77	.49	90.0	72.0	29
MEAN	9.1	20.0	-.18	.53	1.01	-.1	.99	-.1				73.6	72.5	
P.SD	1.9	.0	.53	.02	.36	1.5	.48	1.1				13.3	3.0	

Figure 8. Person Fit Order

The Person Fit Order table in Figure 6 is used to evaluate the fit of students' response patterns to the ideal Rasch model. This is to ascertain whether students are consistently answering the items by their actual abilities. Model inconsistency is indicative of a distorted response set pattern brought about by inconsistency in reasoning, random errors, or suspected cheating [13]. Based on the analysis results, there were students with values outside of Rasch model criteria, namely Mean Square Outfit (MNSQ) outside the range 0.5 to 1.5, Z-Standardised (ZSTD) outside -2.0 to +2.0, and Point Measure Correlation (Pt Mean Corr) values less than 0.4. Student with entry number 41 had Outfit MNSQ value 2.69 and ZSTD 2.28 and Pt Mean Corr -0.25, indicating a very bad fit to the model. On the other hand, On the contrary, students 33 and 35 also had high MNSQ, ZSTD values, and low or negative correlations- this inconsistency suggests a possibility that the participants were not answering the items as expected for their ability level, possibly due to a lack of insight about the items or due to other extraneous factors like guessing or cheating. Thus, analysis of this person-fit order is a constructive exercise in revealing the quality of participant response data and the validity of measures of ability under Rasch-based assessment.

```
GUTTMAN SCALOGRAM OF RESPONSES:
Person |Item
       | 11 1 11 11   2 1 11
       |14075939612245038876
       |--------------------
    40 +11101111110001111010  40
    28 +11111111011110000000  28
    32 +11111010001110100011  32
    34 +11011011100001101011  34
    38 +11100111100001111010  38
    56 +11110011101001101010  56
     2 +11111100001110001100  02
     5 +11111111011100000000  05
    11 +11111111001110000000  11
    14 +11101011001010100101  14
    20 +11110101001110101000  20
    39 +11111000100001110110  39
    49 +11000011100111101010  49
    52 +11100111100001111000  52
     9 +11111111010100000000  09
    16 +11111111001100000000  16
    21 +11111101001010100000  21
    23 +11111110001110000000  23
    35 +10011100110011010100  35
    42 +11100101100101101000  42
    54 +11001100100111011000  54
     1 +11111111001000000000  01
     4 +11111011001010000000  04
     8 +11111110010100000000  08
    10 +11111110001010000000  10
    15 +11111010010000110000  15
    17 +11111010001010000100  17
    19 +11111100001110000000  19
    24 +11111111010000000000  24
    26 +11111110010100000000  26
    27 +11111111010000000000  27
    29 +11111111010000000000  29
    36 +11001101100001100010  36
    50 +11110100100011000001  50
    53 +11100100110001010100  53
     3 +10111111010000000000  03
     7 +11011110001000010000  07
    12 +11110000011100000100  12
    18 +10111111010000000000  18
    22 +11111010010100000000  22
    25 +11110111001000000000  25
    31 +11111011010000000000  31
    43 +11100100100100010100  43
    44 +11100101010000010100  44
    48 +11011010101001000000  48
     6 +11111010010000000000  06
    30 +11110011001000000000  30
    37 +11000000110100010100  37
    45 +01100100100011100000  45
    46 +11001000100000011010  46
    51 +11100000100001011000  51
    55 +11010000100001100001  55
    47 +11000100100001010000  47
    13 +11011000000010000000  13
    33 +00010000000111001000  33
    41 +00000001010010100010  41
       |--------------------
       | 11 1 11 11   2 1 11
       |14075939612245038876
```

Figure 9. Scalogram

The scalogram in Figure 9 shows the inconsistency of the response patterns of some learners to items with varying levels of difficulty. For instance, students 07 and 15 were unable to answer correctly low-difficulty items such as numbers 6 and 0, but were able to answer higher-difficulty items such as numbers 1, 4, or 9. Similarly, student 17 struggled with items 0 and 2 (low difficulty) yet answered correctly on items 1 and 4 (higher difficulty). These anomalies may indicate distraction, disengagement, or inattentiveness during the test. Comparable patterns have been reported in other Rasch studies, where unexpected failures on easy items are linked to random guessing or lack of focus [12], [16]

Furthermore, a study on Newtonian mechanics (bMCU test) found that Rasch-based person-item maps and information curves effectively identified students who performed inconsistently, missing easier items while correctly answering more difficult ones, highlighting potential measurement noise rather than true ability variance. This supports the notion that such irregular response patterns may not reflect conceptual weakness, but rather test-taking behavior, such as fatigue or lack of motivation.

Rasch analysis thus offers tools not only for item and ability estimation, but also for diagnosing anomalies in response behavior that traditional test theory may miss. Future research could employ follow-up interviews or think-aloud protocols to investigate the reasons behind these patterns, in line with instrument development best practices.

D. Conclusion

This study aimed to evaluate students’ conceptual understanding of elasticity and Hooke’s Law using a diagnostic test analyzed with the Rasch model. The test results showed that most students had low to moderate levels of understanding, with only one student categorized as high-ability. Several items, such as S16, S8, and S17, were identified as valid and informative, revealing students' ability differences and common misconceptions. The Rasch model provided detailed information about item difficulty and student ability, which traditional scoring methods often overlook. It also helped identify items that fit well and those that need revision. These findings highlight the usefulness of the Rasch model in developing better assessments and in understanding how students grasp physics concepts.

To improve future instruction and assessment, further research is needed to refine the instrument, expand it to different student populations, and explore its use in various learning contexts. Developing more diverse diagnostic tools, such as two-tier or open-ended items, can also help deepen the evaluation of students' conceptual understanding. This study demonstrates how well-constructed instruments, supported by strong analytical models, can guide better teaching and learning in physics education.

References

[1] P. Phanphech, T. Tanitteerapan, and E. Murphy, "Explaining and Enacting for Conceptual Understanding in Secondary School Physics," *Issues Educ. Res.*, vol. 29, no. 1, pp. 180–204, 2019.

[2] H. J. Banda and J. Nzabahimana, "Effect of Integrating Physics Education Technology Simulations on Students' Conceptual Understanding in Physics: A Review of Literature," *Phys. Rev. Phys. Educ. Res.*, vol. 17, no. 2, pp. 023108-1-023108–18, 2021, doi: 10.1103/PhysRevPhysEducRes.17.023108.

[3] G. H. Wallace, "Improving Spanish Classroom Assessment via Logistic Regression: Lessons from the Rasch Model," *J. Spanish Lang. Teach.*, vol. 7, no. 1, pp. 51–63, 2020, doi: 10.1080/23247797.2020.1771009.

[4] W. L. Nurjanah, I. M. Sari, and D. Saepuzaman, "Analisis Pemahaman Konsep Peserta Didik Pada Topik Perambatan Kalor Menggunakan Rasch Model," *Didakt. J. Kependidikan*, vol. 13, no. 2, pp. 1361–1374, 2024, doi: 10.58230/27454312.442.

[5] B. Sumintono, "Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: Implementasi Penilaian Formatif (Assessment for Learning)," 2016, *Institut Teknologi Sepuluh November, Surabaya*. [Online]. Available: http://eprints.um.edu.my/15876/1/ITS_rasch_model_asesment_for_learning.pdf

[6] R. M. A. Al Ali and R. T. Shehab, "Psychometric Properties of Social Perception of Mathematics: Rasch Model Analysis," *Int. Educ. Stud.*, vol. 13, no. 12, pp. 102–110, 2020, doi: 10.5539/ies.v13n12p102.

[7] J. D. Bransford, A. L. Brown, and R. R. Cocking, *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: National Academy Press, 2000.

[8] D. Hestenes, M. Wells, and G. Swackhamer, "Force Concept Inventory," *Phys. Teach.*, vol. 30, pp. 141–158, 1992, doi: 10.1119/1.2343497.

[9] Z. Zacharia and O. R. Anderson, "The Effects of an Interactive Computer-based Simulation Prior to Performing a Laboratory Inquiry-based Experiment on

Students' Conceptual Understanding of Physics," *Am. J. Phys.*, vol. 71, no. 6, pp. 618–629, Jun. 2003, doi: 10.1119/1.1566427.
[10] T. M. Haladyna, S. M. Downing, and M. C. Rodriguez, "A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment," *Appl. Meas. Educ.*, vol. 15, no. 3, pp. 309–333, Jul. 2002, doi: 10.1207/S15324818AME1503_5.
[11] M. Tarrant and J. Ware, "A Comparison of the Psychometric Properties of Three- and Four-option Multiple-choice Questions in Nursing Assessments," *Nurse Educ. Today*, vol. 30, no. 6, pp. 539–543, 2010, doi: 10.1016/j.nedt.2009.11.002.
[12] W. J. Boone, J. R. Staver, and M. S. Yale, *Rasch Analysis in the Human Sciences*. Springer, 2014.
[13] B. Sumintono and W. Widhiarso, *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata, 2015.
[14] Z. Arsiah *et al.*, "Profil Satuan Pendidikan dengan Capaian AKM Tinggi pada Jenjang SMA/SMK/MA/Sederajat," Jakarta, 2024.
[15] H. Kurli, N. Haliqah, and A. Anekawati, "Analisis Kemampuan Kognitif Siswa pada Mata Pelajaran IPA SMP Menggunakan Rasch Model," in *Prosiding Webinar Nasional Penelitian dan Pengabdian kepada Masyarakat*, Sumenep: Universitas Wiraraja, 2021.
[16] T. G. Bond and C. M. Fox, *Applying the Rasch model: Fundamental Measurement in the Human Sciences*. Routledge/Taylor & Francis Group, 2015.