Thabiea : Journal of Natural Science Teaching
Vol. 6(1), pp. 25-36, 2023
Available online at
http://journal.iainkudus.ac.id/index.php/Thabiea
p-issn: 2580-8474, e-issn: 2655-898X

Rasch model analysis of physics test of HOTS on the topic of elasticity and Hooke's law

Daimul Hasanah[1*], Joko Purwanto[2]
[1]*Universitas Sarjanawiyata Tamansiswa, Jl. Batikan UH-III/1043, 55167, Indonesia*
[2]*Universitas Islam Negeri Sunan Kalijaga Yogyakarta, Jl. Laksda Adisucipto, 55281, Indonesia*
[*]*Correspondence: daimul_hasanah@ustjogja.ac.id*

Keywords:
Elasticity and-Hooke's Law;
HOTS;
Rasch Model;

Abstract

This study aims to measure the quality of the physics learning outcomes test instrument based on high order thinking skills (HOTS) on Elasticity and Hooke's Law and to classify students based on their ability to work on test items. This study uses a quantitative approach. The research was conducted at Sedayu 1st Senior High School with 60 students at eleventh grade in science program as test subjects in the first semester of the 2022/2023 academic year. The data collection technique is done by testing technique. The instrument used is in the form of multiple-choice test questions with 5 alternative answers. The number of test items is 30 test items. The results showed that: first, 27 test items fit the criteria for acceptance of good test items, which can then be used to measure student learning outcomes for physics subjects in Elasticity and Hooke's Law. Three test items are not by the acceptance criteria for good test items, so they need to be revised or replaced with other test items: the competency achievement indicators and the question indicators that are the same as the three test items. Second, the abilities of students who took the test were grouped into 3 groups, namely high, moderate, and low abilities. Students who have high category abilities in terms of working on HOTS-based physics test items are 18.33%. Students with moderate abilities in working on HOTS-based physics tests on Elasticity and Hooke's Law, namely as much as 65.00%, while students with low abilities are as many as 16.67%.

To cite this article:
Hasanah, D., Purwanto, J. (2023). Rasch model analysis of physics test of HOTS on the topic of elasticity and Hooke's law. *Thabiea : Journal of Natural Science Teaching*, 6(1), 25-36.

Introduction

Educational assessment is an essential part of an educational activity process (Sumintono, 2016). The learning process in the classroom always involves educational assessment as one of the most important things to do. Educational assessment is one thing that is very important because, from the educational assessment process, it can be known with certainty whether the learning objectives have been achieved or not. In addition, student learning outcomes can also be used by teachers to: a) determine students' abilities relative to other students in the same test; b) shows the development of students' abilities towards specific knowledge or skills in a certain period; c) show empirical evidence about students' understanding of learning materials; and d) predict student performance in the future (Sumintono, 2016). Almost all tests or assessments conducted in class generally use a scoring approach to describe student learning outcomes. At the same time, there is an unavoidable drawback of using this approach which usually does not support effective feedback.

The classical scoring theory is one of many approaches to educational assessment and psychometrics. There are several other approaches which are alternatives to the classical theoretical approach. Basically, the use of raw scores as a measure of achievement has several drawbacks, including: a) raw scores are basically not measurement results: raw scores are more precisely the number of correct answers to the questions students worked on; b) raw score is initial information: raw score is also usually expressed in percentage (%) which is nothing but a summary of data in the form of numbers, but does not provide data from a measurement; c) the raw score has a weak quantitative meaning: the quantitative meaning of the raw score obtained will be different, depending on the number of questions (items), while the percentage of correct answers always depends on the level of difficulty of the questions (items); d) the raw score does not indicate a person's ability to perform a particular task: the raw score also does not explain much about the level of difficulty of the items (items); and e) the raw score and the percentage of correct answers are not always linear: in a test that is linear, students who have a score of 15 (scale 0 to 100) always have higher abilities than those who have a score of 10, but empirically sometimes both may have the same capabilities (Sumintono, 2016).

Therefore, to overcome the limitations and weaknesses of the classical test theory, other alternatives are needed in analyzing student test results in the form of raw scores. Then, the lack of classical test theory (CTT) is corrected by item response theory (IRT) with various variations of its logistic parameters (PL) (Ekstrand et al., 2022; Putra et al., 2021). One form of the logistic parameter is 1 PL which is developed into the Rasch Model (McArthur, 1987). Unlike CTT, which always depends on scores, IRT does not depend on a particular sample size and the abilities of students (people) involved in the exam or test (Muslihin et al., 2022).

George Rasch developed an analytical model from item response theory (IRT) in the 1960s. The analysis model is usually called 1 PL (one logistic parameter) (Sumintono & Widhiarso, 2015). This mathematical model was later popularized by Benjamin Wright (Linacre, 2017). Through raw data in the form of dichotomous data (true and false) that indicate or represent student abilities, Rasch formulates this into a model that connects students with items (Sumintono & Widhiarso, 2015). In the context of the Rasch Model, the scoring pattern that tends to stick to classical test theory (CTT) is a measurement whose results depend on who is being measured (test-dependent scoring). At the same time, what must be done in quantitative research for educational assessment is an objective measurement (objective measurement). The concept of accurate measurement in social sciences and educational assessment must meet five criteria, including: a) providing a linear measure with equal intervals; b) carrying out an appropriate estimation process; c) finding inappropriate items (misfits) or unusual (outliers); d) address missing data; and e) produce replicable measurements (independent of the parameters studied) (Sumintono, 2016). Of the five criteria, only the Rasch Model can meet these five criteria. Thus, the quality of measurements in educational assessments made using the Rasch Model will have the same quality as measurements made in the physical dimension in physics (e.g. measuring the length of a pencil using a ruler, measuring a child’s body mass using a scale).

In this study, the results of the analysis of physics test data based on high order thinking skills (HOTS) were described in several students. The analysis was carried out by applying the Rasch model using Winsteps software. Compared to CTT, the IRT approach is rarely used in the education sector, especially to analyze student test results to measure student learning outcomes. The IRT approach can provide more comprehensive, consistent, and accurate analytical results compared to the CTT. Therefore, this study seeks to present more comprehensive test results by utilizing the IRT approach through the Rasch Model. Software for the Rasch Model analysis is also easy to obtain and operate, so the Rasch model analysis needs to be applied to teachers to assist teachers in presenting more comprehensive test results.

Previous researchers have carried out research on the Rasch Model (Asriadi & Hadi, 2021; Martinková & Zvára, 2007; Muslihin et al., 2022; Susongko, 2016; Tarigan et al., 2022). This study aims to determine the validity and reliability of a test using the IRT approach through the Rasch Model compared to the CTT approach. The results of this study only discuss the quality of the test in terms of validity and reliability but have not yet discussed the respondents' ability to the test, both comprehensively and individually. Research on the Rasch Model with the theme of HOTS has also been carried out by previous researchers (Munali et al., 2021; Nirwana et al., 2019), but previous studies have not fully described the HOTS indicator. Research on Rasch Model analysis carried out in physics education includes topics on vectors (Susac et al., 2018), static and dynamic fluids (Angraeni et al., 2020), energy (Yusup, 2021), and kinematics (Purwana et al., 2020). The topic that discusses elasticity and Hooke's law has never been studied. This study complements the shortcomings of several previous studies, including the theme of the HOTS test instrument, which has been equipped with HOTS indicators on the topic of elasticity and Hooke's law. Next, the test results were analyzed using the IRT approach through the Rasch Model analysis.

This research aimed to analyze HOTS instruments for physics subjects on Elasticity and Hooke's Law using the Rasch Model (Asriadi & Hadi, 2021; Ibnu et al., 2019; Munali et al., 2021; Nirwana et al., 2019; Planinic et al., 2019; Santos et al., 2016; Susongko, 2016; Tarigan et al., 2022). The results of the analysis of the HOTS test instrument using the Rasch Model will be described regarding: a) the quality of the HOTS-based test instrument for physics subjects on Elasticity and Hooke's Law; and b) the student's ability relative to other students in the same test (Kurniawan & Andriyani, 2018).

Method

This research was conducted using a quantitative approach. The test subjects were 60 students at eleventh grade in science program at Sedayu 1st Senior High School, Yogyakarta. The data collection technique used in this study is the test technique. The test technique measures the quality of HOTS-based test instruments in physics on the topic of Elasticity and Hooke's Law. It describes student abilities relative to other students on the same test. The instrument used is in the form of a multiple-choice test with five alternative answers. The number of test items tested was 30 test items. The test items measure high order thinking skills (HOTS). The HOTS test item indicators are described in Table 1.

Table 1. HOTS-Based Physics Test Instrument Indicators

Number	Indicators of Competence Achievement	Topic and Question Indicator	Cognitive Level	Question Number
1.	Analyze characteristics elastic thing and no elastic.	Topic: Stress and Strain Students can analyze: the point of maximum stress in the conditions carried out; factors that affect the value of the elasticity of objects; the elastic limit point of the object on the graph.	C4	1, 4, 2
		Topic: Young's Modulus Students can analyze materials that have the smallest Young's modulus.	C4	8
2.	Conclude condition elasticity objects from data which exists.	Topic: Stress and Strain Students can conclude: objects that have the most elastic properties; condition of the elastic object when it is under certain conditions.	C5	3, 6
		Topic: Young's Modulus Students can conclude: which elastic material has the largest Young's modulus value; the body conditions match the graph between Young's modulus and strain.	C5	5, 11
		Topic: Potential Energy Students are able to conclude: the right statement; graph of the force against the appropriate length mining to produce the appropriate potential energy.	C5	14, 13
3.	Compare a value elasticity a number of object.	Topic: Young’s Modulus Students can compare: Young's modulus between objects A and B; Young's modulus of both wires; elastic stretch value.	C4	9, 10, 12
		Topic: Potential Energy Students can compare the potential energy of the two rubbers.	C4	15
4.	Interpret related chart elasticity properties object.	Topic: Hooke’s Law Students can interpret the relationship between force and length increase.	C5	18
5.	Analyze elasticity value of object.	Topic: Potential Energy Students can analyze the velocity of the ball when it hits the end of the spring.	C4	17
		Topic: Young's Modulus Students can analyze the resulting strain.	C4	7
6.	Comparing springs series arrangement, parallel, and mixture.	Topic: Hooke's Law Students can compare: the increase in length of the two sets of springs; increase in the length of the two sets of springs.	C4	19, 27
7.	Conclude spring condition series arrangement, parallel, and mixture.	Topic: Hooke’s Law Students can conclude: the right spring constant value; the point that has the smallest spring constant; the condition of the spring when it has loads with different masses.	C5	26, 23, 28
		Topic: Potential Energy Students can infer the condition of a spring when it has a different potential energy.	C5	16
8.	Analyze arrangement spring series, parallel, and mixed.	Topic: Hooke's Law Students can analyze: spring height after being given liquid; trampoline strain to be generated; rock masses that match the conditions of the catapult; increase in the total length of the spring; mass in the spring series so that it has the same length increase; what is the weight of the load W; proper spring arrangement.	C4	20, 21, 22, 24, 25, 29, 30

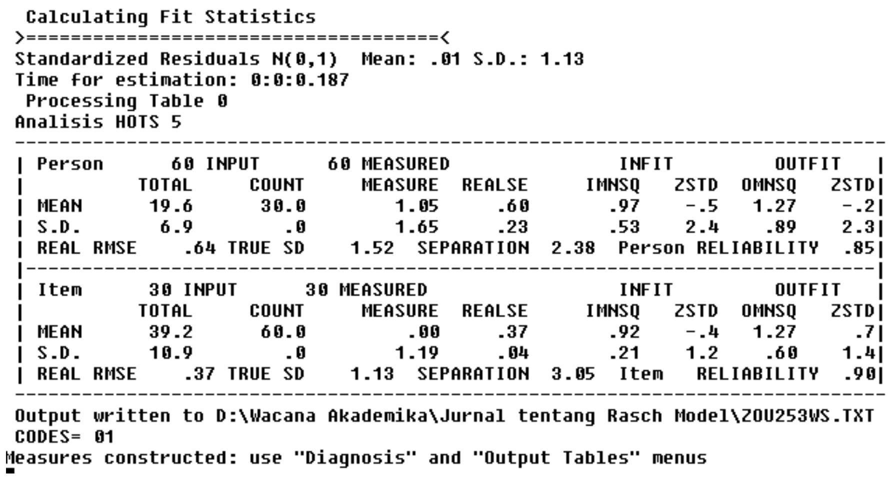
The HOTS test items were tested on 60 students at eleventh grade in science program at Sedayu 1st Senior High School, Yogyakarta. The data from the HOTS item test results that have been obtained are then analyzed. The approach used to analyze the HOTS item test result data is no longer a classical test theory approach (Ding & Beichner, 2009) but a modern data analysis approach (Bond & Fox, 2007). The approach referred to is the Rasch model measurement or Rasch modelling measurement (Sumintono & Widhiarso, 2015). The Rasch model is an essential psychometric tool when conducting science education research using multiple-choice tests (Boone & Scantlebury, 2006). In educational assessment, a different approach using raw scores is needed. This is done to provide more complete information about students' abilities; at the same time, it can also determine the quality of a given HOTS item test (Susac et al., 2018). One of the main goals is to produce a measurement scale with equal intervals, which can provide accurate information about test takers and the quality of the items worked on. The software used to analyze the HOTS item test results using the Rasch model is Winsteps (Linacre, 2017). The criteria used to check the suitability of inappropriate HOTS test items (outliers or misfits) are presented in Table 2 (Sumintono & Widhiarso, 2015).

Table 2. Areas of Acceptance of Fit Items

Number	Description	Areas of Acceptance
1.	Value of *Outfit Mean Square* (MNSQ)	0,5 < MNSQ < 1,5
2.	Value of *Outfit Z-Standard* (ZSTD)	-2,0 < ZSTD < +2,0
3.	Value of *Point Measure Correlation*	0,4 < *Pt. Measure Corr.* < 0,85

Results and Discussion

The results of research data analysis through the Rasch Model using the Winsteps computer application are presented in Figure 1.



```
 Calculating Fit Statistics
>====================================<
Standardized Residuals N(0,1)  Mean: .01 S.D.: 1.13
Time for estimation: 0:0:0.187
 Processing Table 0
Analisis HOTS 5
-------------------------------------------------------------------------------
| Person      60 INPUT      60 MEASURED              INFIT         OUTFIT    |
|          TOTAL      COUNT      MEASURE  REALSE     IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN      19.6       30.0         1.05     .60       .97    -.5   1.27    -.2|
| S.D.       6.9         .0         1.65     .23       .53    2.4    .89    2.3|
| REAL RMSE    .64 TRUE SD    1.52  SEPARATION  2.38  Person RELIABILITY  .85|
|-----------------------------------------------------------------------------|
| Item      30 INPUT      30 MEASURED                INFIT         OUTFIT    |
|          TOTAL      COUNT      MEASURE  REALSE     IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN      39.2       60.0          .00     .37       .92    -.4   1.27     .7|
| S.D.      10.9         .0         1.19     .04       .21    1.2    .60    1.4|
| REAL RMSE    .37 TRUE SD    1.13  SEPARATION  3.05  Item   RELIABILITY  .90|
-------------------------------------------------------------------------------
Output written to D:\Wacana Akademika\Jurnal tentang Rasch Model\ZOU253WS.TXT
CODES= 01
Measures constructed: use "Diagnosis" and "Output Tables" menus
```

Figure 1. Results of Person & Item Reliability Analysis

Based on Figure 1, the personal reliability value of the students involved in measuring the test items' quality obtained a value of 0.85. In contrast, the value of the item reliability test got a value of 0.90 (Martinková & Zvára, 2007).

The results of the analysis of HOTS-based test instrument item quality data for physics subjects on Elasticity and Hooke's Law through Rasch Model analysis using the Winsteps computer application are presented in Figure 2.

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
10	21	60	1.93	.33	.92	-.4	1.33	1.0	.54	.54	81.4	77.2	S10
24	21	60	1.93	.33	.92	-.4	1.47	1.3	.53	.54	84.7	77.2	S24
20	23	60	1.72	.32	.74	-1.9	.69	-1.0	.66	.54	86.4	75.9	S20
11	24	60	1.62	.32	.83	-1.2	1.02	.2	.60	.54	84.7	75.3	S11
1	25	60	1.52	.31	.76	-1.9	.92	-.2	.64	.54	91.5	74.7	S1
2	26	60	1.43	.31	.83	-1.3	.97	.0	.61	.54	86.4	74.3	S2
6	26	60	1.43	.31	1.01	.1	1.73	2.1	.49	.54	83.1	74.3	S6
9	26	60	1.43	.31	.87	-1.0	1.21	.8	.58	.54	86.4	74.3	S9
8	27	60	1.33	.31	1.13	1.0	2.04	2.9	.42	.54	78.0	73.8	S8
3	29	60	1.14	.31	1.20	1.6	2.25	3.5	.36	.54	81.4	72.6	S3
7	32	60	.85	.31	1.48	3.6	2.95	4.9	.20	.54	78.0	70.8	S7
5	43	60	-.29	.34	.91	-.4	1.18	.6	.52	.52	86.4	80.9	S5
18	43	60	-.29	.34	1.10	.6	1.43	1.1	.43	.52	83.1	80.9	S18
14	44	60	-.41	.35	.88	-.6	1.13	.4	.53	.51	88.1	81.8	S14
22	44	60	-.41	.35	.87	-.6	.80	-.3	.57	.51	84.7	81.8	S22
25	45	60	-.53	.36	1.09	.5	2.37	2.3	.39	.51	84.7	82.8	S25
12	46	60	-.66	.37	.75	-1.2	.61	-.7	.62	.50	89.8	83.6	S12
23	46	60	-.66	.37	.87	-.5	1.90	1.6	.49	.50	89.8	83.6	S23
28	46	60	-.66	.37	.56	-2.3	.48	-1.2	.69	.50	93.2	83.6	S28
29	46	60	-.66	.37	.77	-1.1	.81	-.3	.59	.50	89.8	83.6	S29
16	47	60	-.80	.38	.52	-2.5	.56	-.8	.69	.50	94.9	84.4	S16
21	47	60	-.80	.38	.99	.0	1.10	.4	.48	.50	84.7	84.4	S21
17	48	60	-.95	.39	.88	-.4	.92	.1	.52	.49	86.4	85.0	S17
27	49	60	-1.10	.40	.86	-.5	.79	-.1	.53	.48	89.8	85.7	S27
19	50	60	-1.26	.41	.81	-.7	1.43	.8	.49	.47	93.2	86.4	S19
26	50	60	-1.26	.41	1.16	.7	1.52	.9	.36	.47	83.1	86.4	S26
30	50	60	-1.26	.41	1.28	1.1	1.25	.6	.34	.47	83.1	86.4	S30
4	51	60	-1.44	.42	1.17	.7	1.94	1.2	.34	.45	81.4	87.0	S4
13	51	60	-1.44	.42	.67	-1.3	.36	-.9	.61	.45	94.9	87.0	S13
15	51	60	-1.44	.42	.76	-.9	.91	.1	.53	.45	91.5	87.0	S15

Figure 2. Item Analysis Results

If the results of the analysis of the test item data in Figure 2 are confirmed with the acceptance criteria for fit test items in Table 2, the result is that there are 3 test items that do not match the acceptance criteria for fit test items (Table 2). The three test items include test item number 3 (S3), test item number 7 (S7), and test item number 25 (S25). The three test items are not included in the criteria for acceptance of fit test items because all three have outfit mean square (MNSQ) values, outfit Z-standard (ZSTD) values, and point measure correlation (*Pt. Measure Corr.*) values that do not match/do not match with the value of acceptance of fit test items. Because the three test items (S3, S7, and S25) are not fit, the three test items need to be changed/revised with changes according to needs. If stated in percentage numbers, then the number of test items that fit is 90%, while the number of outfitted test items is 10%.

From the results of the data analysis in Figure 2, the HOTS test items can be identified by the difficulty level of the questions with the criteria of difficult, very difficult, easy, or very easy. Identifying the HOTS test item groups is based on the standard deviation (SD) values resulting from the Rasch Model analysis. The standard deviation value obtained is 1.19. HOTS test items are included in the difficult criteria if the logit value is 0.0 logit + 1SD, while the HOTS test items are included in the very difficult criteria if the logit value is greater than +1SD. Meanwhile, HOTS test items can be identified as a group of items with a difficulty level of questions with easy criteria if the logit value is 0.0 logit – 1SD, while the criteria are very easy if the logit value is smaller than –1SD. Identification of item groups based on the difficulty level of the items is presented in Table 3.

Table 3. Identification of Item Groups Based on Item Difficulty Level

Logit Value	Items Criteria	Number of Test Item	Amount	Percentage
> 1.19	Very difficult	S10, S24, S20, S11, S1, S2, S6, S9, S8	9	30.00%
0.01 – (+1.19)	Difficult	S3, S7	2	6.67%
(-1.19) – 0.00	Easy	S5, S18, S14, S22, S25, S12, S23, S28, S29, S16, S21, S17, S27	13	43.33%
< (-1.19)	Very Easy	S19, S26, S30, S4, S13, S15	6	20.00%
Amount				**100%**

The results of analysing students' ability data working on HOTS-based test instrument items for physics subjects on Elasticity and Hooke's Law through Rasch Model analysis using the Winsteps computer application are presented in Figures 3.a and 3.b. From Figure 3.a and Figure 3.b, the results of the logit person value are obtained in the "measure" column, which can identify students' abilities in working on the HOTS test items. Based on the results of data analysis using the Rasch Model, the results show that the average logit person ($\bar{x}$) ability of students is 1.05, while the standard deviation (SD) value of the student's ability is 1.65. Furthermore, information about the average logit ($\bar{x}$) value of students' abilities and the value of the standard deviation (SD) is used to classify students' abilities. In this study, students' abilities in working on HOTS test items were grouped into 3 criteria, namely high, moderate, and low abilities. Students are categorized as having a high capability in working on HOTS test items if their logit value is greater than or greater than +2.70 logit. Students are categorized as having moderate ability if their logit values are in the interval up to or within the interval -0.60 logit to 2.70 logit, while the student's ability is included in the low ability criteria if the logit value is less than or less than the logit value -0.6 logit. Identification of student groups based on their abilities is presented in Table 4.

Table 4. Identification of Student Groups Based on Their Abilities

Logit Value	Criteria for Student Abilities	Student Number	Amount	Percentage
$> \bar{x} + 1SD$ (> +2.70 *logit*)	High	48, 38, 39, 40, 41, 37, 44, 49, 51, 52, 59.	11	18.33%
$\bar{x} - 1SD$ until $\bar{x} + 1SD$ (-0.60 *logit* until 2.70 *logit*)	Moderate	47, 50, 43, 53, 55, 46, 54, 56, 09, 36, 45, 15, 16, 17, 18, 19, 20, 21, 23, 24, 08, 10, 11, 12, 13, 14, 22, 26, 28, 30, 31, 32, 33, 35, 57, 29, 58, 60, 25.	39	65.00%
$< \bar{x} - 1SD$ (< -0.6 *logit*)	Low	04, 34, 02, 03, 06, 07, 27, 42, 01, 05.	10	16.67%
Amount				**100%**

TABLE 17.1 Analisis HOTS 5 ZOU228WS.TXT May 12 13:32 2023
INPUT: 60 Person 30 Item REPORTED: 60 Person 30 Item 2 CATS WINSTEPS 3.73

Person: REAL SEP.: 2.38 REL.: .85 ... Item: REAL SEP.: 3.05 REL.: .90

Person STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Person
48	30	30	5.26	1.84	MAXIMUM MEASURE				.00	.00	100.0	100.0	48P
38	29	30	4.00	1.03	1.06	.4	.81	.3	.13	.19	96.7	96.6	38P
39	29	30	4.00	1.03	1.06	.4	.81	.3	.13	.19	96.7	96.6	39L
40	29	30	4.00	1.03	1.06	.4	.81	.3	.13	.19	96.7	96.6	40P
41	29	30	4.00	1.03	1.06	.4	.81	.3	.13	.19	96.7	96.6	41L
37	28	30	3.23	.76	1.04	.3	1.63	.8	.16	.27	93.3	93.3	37L
44	28	30	3.23	.76	1.13	.4	1.71	.9	.10	.27	93.3	93.3	44P
49	28	30	3.23	.76	1.04	.3	1.83	1.0	.14	.27	93.3	93.3	49L
51	27	30	2.75	.64	1.13	.4	2.54	1.5	.13	.32	90.0	89.9	51L
52	27	30	2.75	.64	1.26	.7	1.94	1.1	.04	.32	90.0	89.9	52P
59	27	30	2.75	.64	.93	.0	.50	-.4	.41	.32	90.0	89.9	59L
47	26	30	2.37	.58	1.41	1.2	2.60	1.8	-.05	.37	86.7	86.6	47L
50	26	30	2.37	.58	1.16	.6	1.30	.6	.22	.37	86.7	86.6	50P
43	25	30	2.06	.54	1.22	.8	2.08	1.5	.16	.41	83.3	83.3	43L
53	25	30	2.06	.54	1.51	1.7	2.52	2.0	-.08	.41	83.3	83.3	53L
55	25	30	2.06	.54	1.57	1.8	2.67	2.1	-.13	.41	83.3	83.3	55L
46	24	30	1.79	.51	1.46	1.7	2.12	1.8	.03	.44	80.0	80.4	46P
54	24	30	1.79	.51	1.33	1.2	1.66	1.2	.17	.44	80.0	80.4	54P
56	24	30	1.79	.51	1.66	2.2	2.39	2.1	-.12	.44	73.3	80.4	56P
9	23	30	1.54	.49	.63	-1.7	.40	-1.6	.74	.46	90.0	78.3	09L
36	23	30	1.54	.49	1.82	2.8	2.29	2.2	-.15	.46	63.3	78.3	36P
45	23	30	1.54	.49	1.57	2.1	2.29	2.2	-.03	.46	76.7	78.3	45L
15	20	30	.89	.45	.40	-3.3	.33	-2.9	.92	.51	100.0	77.1	15L
16	20	30	.89	.45	.40	-3.3	.33	-2.9	.92	.51	100.0	77.1	16P
17	20	30	.89	.45	.40	-3.3	.33	-2.9	.92	.51	100.0	77.1	17L
18	20	30	.89	.45	.40	-3.3	.33	-2.9	.92	.51	100.0	77.1	18P

Figure 3.a Results of Person Analysis (26 students)

19	20	30	.89	.45	.40	-3.3	.33	-2.9	.92	.51	100.0	77.1	19L
20	20	30	.89	.45	.40	-3.3	.33	-2.9	.92	.51	100.0	77.1	20P
21	20	30	.89	.45	.40	-3.3	.33	-2.9	.92	.51	100.0	77.1	21L
23	20	30	.89	.45	.40	-3.3	.33	-2.9	.92	.51	100.0	77.1	23L
24	20	30	.89	.45	.40	-3.3	.33	-2.9	.92	.51	100.0	77.1	24P
8	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	08P
10	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	10P
11	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	11L
12	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	12P
13	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	13L
14	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	14P
22	19	30	.69	.44	.70	-1.4	.63	-1.4	.73	.52	86.7	76.9	22P
26	19	30	.69	.44	.70	-1.4	.63	-1.4	.73	.52	86.7	76.9	26P
28	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	28P
30	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	30P
31	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	31L
32	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	32P
33	19	30	.69	.44	.36	-3.7	.32	-3.3	.95	.52	100.0	76.9	33L
35	19	30	.69	.44	2.25	4.1	2.36	3.6	-.34	.52	33.3	76.9	35L
57	19	30	.69	.44	1.66	2.5	1.76	2.3	.06	.52	53.3	76.9	57L
29	18	30	.50	.44	.49	-2.8	.54	-2.0	.85	.52	96.7	76.5	29L
58	18	30	.50	.44	2.19	4.0	2.22	3.6	-.29	.52	36.7	76.5	58P
60	17	30	.31	.43	2.08	3.9	2.20	3.8	-.24	.52	40.0	75.7	60P
25	13	30	-.41	.42	1.49	2.5	1.79	2.5	.09	.48	53.3	70.6	25L
4	10	30	-.95	.43	1.23	1.4	1.86	2.1	.18	.43	76.7	70.8	04P
34	9	30	-1.13	.44	1.21	1.3	1.55	1.3	.20	.41	76.7	72.1	34P
2	7	30	-1.54	.46	1.38	1.8	3.27	3.1	-.14	.36	76.7	76.6	02P
3	7	30	-1.54	.46	1.35	1.6	2.49	2.3	-.06	.36	76.7	76.6	03L
6	7	30	-1.54	.46	1.07	.4	1.75	1.4	.22	.36	76.7	76.6	06P
7	7	30	-1.54	.46	1.25	1.2	1.50	1.0	.13	.36	76.7	76.6	07L
27	7	30	-1.54	.46	1.12	.6	1.95	1.7	.19	.36	76.7	76.6	27L
42	5	30	-2.02	.52	1.24	.9	1.81	1.2	.07	.31	83.3	83.3	42P
1	4	30	-2.30	.56	1.13	.5	2.74	1.8	.01	.28	86.7	86.6	01L
5	4	30	-2.30	.56	.98	.1	1.66	1.0	.22	.28	86.7	86.6	05L
MEAN	19.6	30.0	1.05	.55	.97	-.5	1.27	-.2			86.5	80.8	
S.D.	6.9	.0	1.65	.23	.53	2.4	.89	2.3			16.3	6.7	

Figure 3.b Results of Person Analysis (34 students)

From the results of data analysis in Figure 3.a and Figure 3.b, the Outfit Mean Square (MNSQ) values, Outfit Z-Standard (ZSTD) values, and Point Measure Correlation (Pt. Measure Corr.) values are obtained. These values are used as the criteria for person examination. The criteria used for examining persons are the same as those used for examining test items (Table 2). Based on matching between Outfit Mean Square (MNSQ) Values, Outfit Z-Standard (ZSTD) Values, and Point Measure Correlation (Pt. Measure Corr.) Values with the criteria in Table 2, the result was that 33 students did not meet the three criteria in Table 2, while 27 students met the criteria in Table 2 (even though they only met one criterion), hereinafter referred to as person fit. These results can be used to detect students whose answer patterns are not appropriate, meaning that there are discrepancies in the answers given by students based on their abilities compared to the ideal model. These results can also be used by the teacher to determine the consistency of students' thinking and can also be used to find out if there is fraud committed by students.

```
                                 Person - MAP - Item
                                    <more>|<rare>
 5                                      48  +
                                           T|
 4                             38 39 40 41  +
                                  37 44 49  |
 3                                          +
                                  51 52 59  |
                                           S|
                                     47 50  |T
 2                                43 53 55  +  S10   S24
                                  46 54 56  |
                                            |  S20
                                  09 36 45  |  S1    S11
                                            |  S2    S6    S9
                                            |  S8
                                            |S S3
 1                                         M+
                  15 16 17 18 19 20 21 23 24  |  S7
  08 10 11 12 13 14 22 26 28 30 31 32 33 35 57  |
                                     29 58  |
                                        60  |
 0                                          +M
                                            |  S18   S5
                                        25  |  S14   S22
                                           S|  S25
                                            |  S12   S23   S28   S29
                                            |  S16   S21
-1                                      04  +  S17
                                        34  |S S27
                                            |  S19   S26   S30
                                            |  S13   S15   S4
                            02 03 06 07 27  |
-2                                      42  +
                                           T|
                                     01 05  |
                                            |T
```

Figure 4. Variable Maps (Wright Map)

One of the results of data analysis using the Rasch Model is a map that describes the distribution of student abilities and the distribution of item difficulty levels with the same scale (Figure 4). This map is also called the Wright Map (taken from the name of its inventor Benjamin Wright), a comprehensive person-item (Azizah & Wahyuningsih, 2020; Maryati et al., 2019). The Wright map on the left depicts a student's ability relative to other students on the same test. Analysis of students' abilities using the Rasch Model can be done easily because a scale with the same distance is used in this Rasch Model analysis. Furthermore, the scale is called the logit. From Figure 4, one student has a high capability, namely student number 48. The logit value of student number 48 is +5.26 logit. Students with the lowest abilities are students' number 01 and 05, with a logit value of -2.30 logit. When comparing the ability of student number 48 (with a logit value of +5.26 logit) to students number 29 and number 58 (with a logit value of +0.50 logit), it can be explained that the ability of student number 48 is ten times that of students number 29 and number 58, in the context of students can work on HOTS test instruments (Isnani et al., 2019).

The Wright map on the right describes the distribution of logit item values (Figure 4). Items S10 and S24 are test items with the highest difficulty level (logit item value of +1.93), which means that the probability of all students doing the item correctly is very small. Meanwhile, items S4, S13, and S15 are items with a low logit score (-1.44 logit); in this case, almost all students can work on these items correctly. In addition, items S4, S13, and S15 have the same logit value (-1.44 logit), indicating no different level of difficulty.

The results of data analysis describing the distribution of student abilities and the distribution of item difficulty levels can be combined as follows. First, the logit means the item value is always set within 0.0 logit, which represents the initial reference point of the scale. The average person's logit is +0.89 logit, which means above 0.0 logit. This shows that the average student achievement is above the average standard item difficulty level. Second, student number 48 can do all items (30 items) correctly. This is because the ability of student number 48, whose value is +5.26 logit, is higher than the difficulty level of items number 10 and number 24 (S10 and S24), whose logit value is only +1.93 logit. The logit value of the student's ability which is lower than the logit item, means that the probability of being able to do the item with the larger logit correctly is less than 50%. Student number 48 will have no difficulty doing all the items correctly because the difficulty level of all items is below his ability. Third, students with the same logit scores, namely students' numbers 09, 36, and 45 (+1.54 logit), were compared with items that had almost the same logit scores, namely S1 items (+1.52 logit) and S11 items (+1.62 logit), the probability that the three students can do S1 and S11 items correctly is 50%. Students number 09, 36, and 45 can work on items S2, S6, S9, S8, and S3 well because their level of difficulty is lower than their ability, but they may not be able to do items S20, S10, and S24 correctly, which are the difficulty is higher. Fourth, student numbers 02, 03, 06, 07, 27 can be stated not to have the ability to answer almost all items correctly. This is because even the easiest item with a value of -1.44 logit is still higher than its ability (-1.54 logit). Student number 42, whose ability is worth -2.02 logit and student number 01 and 05, whose ability is worth -2.30 logit (Sumintono & Widhiarso, 2015).

Conclusion

Based on data analysis with the Rasch Model, the research results can be described as follows. First, of the 30 HOTS-based physics test items on the subject matter of Elasticity and Hooke's Law that have been tested, there are 27 test items that are accepted as instruments for measuring student physics learning outcomes because the test items are *fit*. Three test items had to be revised or discarded due to *outfit* (not meeting the 3 test item acceptance criteria). Second, of the 60 students who took the test, 11 students (18.33%) had high ability, 39 students

(65.00%) had moderate ability, and 10 students (16.67%) had low ability in terms of doing HOTS-based physics tests on subject matter Elasticity and Hooke's Law.

Credit Authorship Contribution Statement

Daimul Hasanah: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Joko Purwanto**: Software, Resources, Writing – review & editing, Supervision.

References

Asriadi, M., & Hadi, S. (2021). Analysis of the Quality of the Formative Test Items for Physics Learning Using the Rasch Model in the 21st Century Learning. *JIPF (Jurnal Ilmu Pendidikan Fisika)*, *6*(2). https://doi.org/10.26737/jipf.v6i2.2030

Azizah, & Wahyuningsih, S. (2020). Penggunaan Model Rasch Untuk Analisis Instrumen the Use of Rasch Model for Analyzing Test. *J U P I T E K Jurnal Pendidikan Matematika*, *3*(1).

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model : Fundamental Measurement in the HumBond, T. G., & Fox, C. M. (2007). Applying the Rasch Model : Fundamental Measurement in the Human Sciences Second Edition University of Toledo.an Sciences Second Edition University of Toledo*.

Boone, W. J., & Scantlebury, K. (2006). The role of rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education*, *90*(2), 253–269. https://doi.org/10.1002/sce.20106

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics - Physics Education Research*, *5*(2), 1–17. https://doi.org/10.1103/PhysRevSTPER.5.020103

Ekstrand, J., Westergren, A., Årestedt, K., Hellström, A., & Hagell, P. (2022). Transformation of Rasch model logits for enhanced interpretability. *BMC Medical Research Methodology*, *22*(1). https://doi.org/10.1186/s12874-022-01816-1

Ibnu, M., Indriyani, B., Inayatullah, H., & Guntara, Y. (2019). Aplikasi Rasch Model: Pengembangan Instrumen Tes untuk Mengukur Miskonsepsi Mahasiswa. *Prosiding Seminar Nasional Pendidikan FKIP*, *2*(1).

Isnani, I., Utami, W. B., Susongko, P., & Lestiani, H. T. (2019). Estimation of college students' ability on real analysis course using Rasch model. *Research and Evaluation in Education*, *5*(2). https://doi.org/10.21831/reid.v5i2.20924

Kurniawan, U., & Andriyani, K. D. K. (2018). Analisis Soal Pilihan Ganda dengan Rasch Model. *Jurnal Statistika*, *6*(1).

Linacre, J. M. (2017). *Winsteps Rasch measurement computer program*. Winsteps.com.

Martinková, P., & Zvára, K. (2007). Reliability in the rasch model. *Kybernetika*, *43*(3).

Maryati, Prasetyo, Z. K., Wilujeng, I., & Sumintono, B. (2019). Measuring teachers' pedagogical content knowledge using many-facet rasch model. *Cakrawala Pendidikan*, *38*(3). https://doi.org/10.21831/cp.v38i3.26598

McArthur, D. L. (1987). The Rasch Model for Item Analysis. In *Alternative Approaches to the Assessment of Achievement*. https://doi.org/10.1007/978-94-009-3257-9_4

Munali, Rahayu, W., & Hariwibowo, H. (2021). Analysis Of the Characteristics Of HOTS-Based Assessment for Learning Items Using the Rasch Model. *Review of International Geographical Education Online*, *11*(5). https://doi.org/10.48047/rigeo.11.05.149

Muslihin, H. Y., Suryana, D., Ahman, Suherman, U., & Dahlan, T. H. (2022). Analysis of the Reliability and Validity of the Self-Determination Questionnaire Using Rasch Model. *International Journal of Instruction*, *15*(2). https://doi.org/10.29333/iji.2022.15212a

Nirwana, N., Rochman, S., & Zukmadini, A. Y. (2019). *An assessment of Higher Order Thinking Skills (HOTS) Based on Rasch Models of Student in Physics Learning*. https://doi.org/10.2991/icetep-18.2019.40

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

Putra, Z. H., Hermita, N., & Alim, J. A. (2021). Analisis Pengetahuan Matematika, Didaktika, dan Teknologi Calon Guru Sekolah Dasar Menggunakan Rasch Model. *Mosharafa: Jurnal Pendidikan Matematika*, *10*(3). https://doi.org/10.31980/mosharafa.v10i3.1042

Santos, S., Cadime, I., Viana, F. L., Prieto, G., Chaves-Sousa, S., Spinillo, A. G., & Ribeiro, I. (2016). An application of the Rasch model to reading comprehension measurement. *Psicologia: Reflexao e Critica*, *29*(1). https://doi.org/10.1186/s41155-016-0044-6

Sumintono, B. (2016). Aplikasi Pemodelan Rasch pada Asesmen Pendidikan: Implementasi Penilaian Formatif (assessment for learning). In *Makalah dipresentasikan dalam Kuliah Umum pada Jurusan Statistika, Institut Teknologi Sepuluh November, Surabaya, 17 Maret 2016.* (Issue March, pp. 1–19). http://eprints.um.edu.my/15876/1/ITS_rasch_model_asesment_for_learning.pdf

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

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*, *14*(2), 23101. https://doi.org/10.1103/PhysRevPhysEducRes.14.023101

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

Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D. (2022). Analisis Instrumen Tes Menggunakan Rasch Model dan Software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, *16*(2). https://doi.org/10.15294/jipk.v16i2.30530