IJELLACUSH International Journal of Education, Language, Literature, Arts, Culture, and Social Humanities E-ISSN: 2962-8725

Article

Analysis of 4-Level Web-Based Instruments on Dynamic Electricity Material Using the Rasch Model for High School Learners

Moh Dimas Feri Hermansyah[1], Ernasari [2*], Maryani[3]

[1] Jember University, Indonesia; email: mohdimas328@gmail.com
[2] Jember University, Indonesia; email: ernasari.fkip@unej.ac.id
[3] Jember University, Indonesia; email: maryani.fkip@unej.ac.id

* Corresponding Author : Ernasari

Abstract: Web-based 4-tier instrument analysis on dynamic electricity material is one of the diagnostic tests that can determine the extent of understanding of the concepts that students have according to the level of confidence of the answers and the reasons they answer. The purpose of this study was to determine the feasibility of the instrument using Rasch Model analysis. Rasch Model analysis is assisted by Winstep software to analyze the scores generated from test instruments with the aim of knowing MNSQ Outfit, ZSTD Outfit, Item Measure, Item fit order, Item reliability, and Alpha Cornbach. This research is a descriptive type with quantitative descriptive research methods. The sample of this study was taken from the population derived from class XII students with a total of 91 people who have learned the concept of dynamic electricity in Jember city with purposive sampling technique. The instrument used was a web-based Four-tier diagnostic test consisting of 30 questions. The Rasch Model analysis results stated that the person reliability values of 0.76 and 0.78 were in the sufficient category, the item reliability values of 0.93 and 0.94 were in the excellent category, and the Cronbach Alpha Alpha (reliability) value of 0.78 was in the good category, indicating that the interaction between respondents and items was in the good category. This shows that the diagnostic test instrument is feasible, valid, and reliable.

Keywords: Four-Tier Diagnostic Test, Instrumen Feasibility, Rasch Model.

Received: 19 March, 2025
Revised: 17 April, 2025
Accepted: 05 may, 2025
Published: 08 may, 2025
Curr. Ver.: 08 May, 2025

1. Introduction

The development of technology, information and communication in the current era brings enormous changes in various fields, especially education. This makes a challenge to the curriculum to be more sensitive in the process of preparing a strategic educational framework for improving the quality of education. This quality improvement can be done by improving the 21st century skills needed, namely 4C (creative thinking, critical thinking and problem solving, communication, and collaboration) (Astuti et al., 2019). One of the subjects that plays an important role in achieving the goal of improving 21st century skills is physics (Maison et al., 2019).

Physics is one of the subjects that really requires a good understanding of concepts (Giancoli, 2016). Understanding concepts in physics learning is very important, because students are expected to gain the ability to understand events and be able to solve physics problems so that they can develop the abilities they acquire (Mufit et al., 2019). One of the physics materials that require a good understanding of concepts is dynamic electricity.

The four-tier diagnostic test is a multiple-choice diagnostic test instrument that has the advantage of identifying and having the ability to reveal the level of confidence of students regarding how much confidence students have in the answers and reasons for the answers given. The results of this diagnostic test can inform the level of conceptual understanding of students (Anggraini, 2019). The four-tier diagnostic test is one of the diagnostic tests that has advantages compared to other diagnostic tests because it can determine the extent of concept understanding that students have according to the level of confidence in the answers and reasons they answer (Anggraini et al., 2021).

Diagnostic tests using the conventional PBT (Paper Based Test) method still tend to be chosen by many schools today, although it has many disadvantages. The disadvantages of using paper-based exams are the difficulty of modifying questions, limitations in the form of question displays, and the need for more time to process data and exam results. An alternative solution that can be done is to maximize the use of developing technology and communication (ICT) (Martin, 2020). The application of e-diagnostic tests can be in the form of applications or web. The use of e-diagnostic tests can be an alternative to maximize the use of technology because it is more practical. The use of web-based diagnostic tests will produce a concept understanding profile so that students' concept weaknesses are identified and educators can carry out further academic policies. Improvements made by educators are expected to strengthen students in understanding concepts so that learning objectives can be achieved optimally (Arcelay, 2021).

To evaluate whether a measurement instrument is appropriate, the Rasch Model can be used as a reliable analytical method. One of its key advantages lies in its ability to estimate missing data based on consistent response patterns (Suryana et al., 2020). Since it adopts a probabilistic approach, the Rasch Model avoids deterministic assumptions and enables more accurate measurement results. Additionally, it allows researchers to examine how test-takers interact with items and whether the instrument is valid. Key components commonly analyzed through this model include unidimensionality, Wright map, item analysis, respondent ability, and overall instrument quality (Adimayuda et al., 2020).

A test instrument is considered to have good quality if it demonstrates a high level of validity and reliability. High validity and reliability increase the accuracy of the data collected during research (Bodzin et al., 2020). In a study conducted by Azizah and Wahyuningsih, they evaluated the test instruments used to measure students' abilities in actuarial mathematics courses at the Department of Mathematics, State University of Malang. The Rasch Model was employed to identify items that met the fit criteria (Azizah & Wahyuningsih, 2020), with the assistance of the Winsteps software. The analysis revealed that 25 items were aligned with the Rasch Model. Furthermore, Salsabila et al. emphasized that the Rasch Model enables more accurate, objective, and consistent measurement results, as it clearly reflects the interaction between respondents and statement items within the instrument (Fulmer, 2016).

The importance of instrument analysis in learning is very necessary. This is to show that the instrument given to students is feasible and capable of measuring the level of understanding or in accordance with the objectives to be achieved by educators. Therefore, in this study the authors will examine with the title “Analysis of Web-Based 4-Tier Instruments on Dynamic Electricity Material Using the Rasch Model”.

2. Literature Review

2.1. Four-tier Diagnostic Test

A test is a measurement technique in which there are various questions, statements, or a series of tasks that must be done or answered by respondents used to measure skills, knowledge, intelligence, abilities or talents possessed by individuals or groups (Lengkong et al., 2021). Diagnostic tests are conducted by teachers as an initial step in determining the success of students in understanding concepts during the learning process. The test results will provide information about concepts that have not been understood by students (Diani et al., 2019). The test must contain material that is considered difficult by students, but the level of difficulty is adjusted to the ability of students. The function of diagnostic tests is to identify difficulties in understanding the concepts experienced by students to follow up on problems experienced by students with misconceptions (Tumanggor et al., 2020).

The four-tier diagnostic test is an innovation in the development of assessment instruments designed to explore students' conceptual understanding by considering their confidence levels in both the answers chosen and the reasons behind those choices. This test is an extension of the three-tier multiple-choice diagnostic test, adding an element of students' confidence in their chosen answers and the reasons provided. In the first tier, the instrument presents a multiple-choice question with three distractors and one correct answer that students must select. The second tier measures the students' confidence in their chosen answer. The third tier assesses the reason students provide for their answer, which includes three provided reasons and one open-ended reason. In the fourth tier, students are asked to assess their confidence level in the reason they selected. The main advantages of using this four-tier diagnostic test are (Kaniawati et al., 2020):

a. The ability to clearly distinguish students' confidence in their answers and the reasons they choose, allowing for a deeper exploration of students' conceptual understanding.
b. The ability to diagnose students' misconceptions more thoroughly.
c. The capability to identify areas of the material that require further emphasis in teaching.
d. The support it provides in planning better lessons aimed at reducing students' misconceptions.

2.2. Web-based Four-tier Instrumen

Diagnostic test instruments can utilize technology as a support in identifying misconceptions held by students. The steps that need to be taken in developing diagnostic tests with electronic media are the preparation of diagnostic test instruments and having the necessary system characteristics of diagnostic tests with electronic media. Electronic media can be web-based (computer based test) (Arcelay, 2021). In this study, web-based diagnostic tests are media built using HTML and Google Script on the Google Appscript platform. The reasons for using Google Script are 1) accessibility and ease of use as long as it is connected to the internet, 2) integration that is directly connected to other Google services such as Google Sheets or Google Forms, 3) Cost efficient, and 4) fast and flexible development when updates are made. Web-based diagnostic tests have their own advantages compared to manual tests in general. Working on questions is done directly through the web so that data management of test results can be done more quickly and facilitate the evaluation process. In appearance and practicality, web-based diagnostic test instruments can be favored (Istiyono et al., 2023).

2.3. Dynamic Electricity

Dynamic electricity is electricity that is not fixed or can move, this electrical displacement is called an electric current. Electric current is the movement of electric charge in the form of electrons in an electrical circuit within a certain time due to electric voltage (Haertel, 1984). Electric current is created from a constant flow of negative charges flowing from the negative pole to the positive pole, from high voltage to low voltage and from a source of potential difference (voltage). Electric current is divided into two, namely direct current (DC) whose electric flow is always fixed and constant over time and only has one positive to negative direction, and alternating electric current (AC) whose electric flow moves back and forth in both direction and magnitude. Sub-discussions in dynamic electricity material are electric current, ohm's law, Kirchoff's Law I and II, series and parallel circuits (Giancoli, 2016).

3. Method

This study aims to describe the feasibility of the developed instrument using a quantitative descriptive method. The research was conducted at a public high school in Jember City during the even semester of the 2024/2025 academic year, involving 91 students from grade XII who have studied the concept of dynamic electricity. The sample was selected using purposive sampling, as no similar research had been conducted in Jember, and no web-based instrument was available to diagnose students' misconceptions.

The instrument used in this study was a web-based four-tier diagnostic test consisting of 30 questions. This instrument was designed to assess students' understanding of the dynamic electricity concept and to evaluate any misconceptions they may have. The instrument's validity was tested using the Rasch Model, analyzed with the help of Winstep software. Winstep was used to evaluate the Outfit MNSQ, Outfit ZSTD, Item Measure, Item fit order, Item reliability, and Alpha Cronbach, which are key indicators in assessing the quality of the test instrument.

4. Results and Discussion

Analysis of the feasibility of test instruments using Rasch Model analysis, which refers to the feasibility of question items, the level of difficulty of the questions, the validity test of the question items, and the reliability of the instruments that have been made. The results of students' answers were then analyzed using the Rasch Model with the following results.

4.1. Summary Statistics

```
     SUMMARY OF 91 MEASURED PERSON
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      13.1      30.0       -.30      .43        .99    -.1   1.03     .1 |
| S.D.       5.1        .0        .92      .05        .21    1.1    .37    1.1 |
| MAX.      25.0      30.0       1.88      .66       1.72    3.0   2.22    3.1 |
| MIN.       3.0      30.0      -2.60      .40        .68   -2.6    .52   -1.8 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .45 TRUE SD     .80  SEPARATION  1.76  PERSON RELIABILITY  .76 |
|MODEL RMSE    .44 TRUE SD     .81  SEPARATION  1.86  PERSON RELIABILITY  .78 |
| S.E. OF PERSON MEAN = .10                                                   |
-------------------------------------------------------------------------------
PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .78

     SUMMARY OF 30 MEASURED ITEM
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      39.8      91.0        .00      .25       1.00    -.2   1.03    -.1 |
| S.D.      16.5        .0        .99      .03        .13    1.2    .30    1.4 |
| MAX.      81.0      91.0       1.66      .35       1.38    2.8   2.05    3.7 |
| MIN.      14.0      91.0      -2.70      .23        .77   -2.8    .71   -2.6 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .26 TRUE SD     .95  SEPARATION  3.71  ITEM   RELIABILITY  .93 |
|MODEL RMSE    .25 TRUE SD     .96  SEPARATION  3.82  ITEM   RELIABILITY  .94 |
| S.E. OF ITEM MEAN = .18                                                     |
-------------------------------------------------------------------------------
```

Figure 1. Summary statistics.

The results of summary statistics are used to see three types of criteria, namely, person reliability, item reliability and Cronbach Alpha. These three types of criteria are used to assess the extent to which participants' answers to question items match the expected criteria. The criteria for summary statistics results are in accordance with the provisions in Tables 1 and 2 (Dewi et al., 2021).

Table 1. Person reliability and Item reliability categories

Person reliability and Item reliability values Classification	Classification
≤ 0,67	Weak
0,67 < x ≤ 0,80	Fair
0,80 < x ≤ 0,90	Good
0,90 < x ≤ 0,94	Very good
> 0,95	Special

Table 2. Category Cronbach Alpha value (reliability) of question items

Cronbach Alpha (reliability) value	Classification
≤ 0,50	Weak
0,50 < x ≤ 0,60	Bad
0,60 < x ≤ 0,70	Fair
0,70 < x ≤ 0,80	Good
> 0,80	Very good

The summary statistics shown in Figure 1 indicate that the person reliability values (0.76 and 0.78) fall within the sufficient category, while the item reliability values (0.93 and 0.94) are in the excellent category. Person reliability reflects the consistency of respondents' answers, while item reliability reflects the quality of the instrument items. Although the consistency of students' answers is relatively weak, the quality of the items is very high. The Cronbach Alpha value of 0.78 suggests a good interaction between respondents and items. The Rasch model provides more detailed information on the relationship between the item characteristics and the individuals responding to them (Soeharto, 2021). Koçak emphasizes that reliability also indicates the consistency of statement items within an instrument. These high reliability test results indicate that the instrument is dependable for assessing students' ability dimensions (Koçak, 2020).

4.2. Item Fit Order

The Item Fit Order value reflects the degree of alignment of the items used in the instrument, aiming to assess whether the items are capable of distinguishing students based on their ability levels. In this context, the expected value for Outfit Mean Square (MNSQ) falls within the range of 0.5 < MNSQ < 1.5, which indicates a good fit between the items and respondents. Meanwhile, the Outfit Z-standard (ZSTD) value should fall within the range of -2.0 < ZSTD < +2.0, suggesting that the data does not contain significant deviations. Furthermore, the Point Measure Correlation (Pt Measure Corr) value is expected to fall between 0.20 < Pt Measure Corr < 0.60, indicating a moderate positive correlation between the items and the scores produced. These three indicators provide valuable insights to ensure that the items in the research instrument effectively capture the differences in students' abilities.

ITEM STATISTICS: MISFIT 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
5	21	91	1.10	.27	1.36	2.4	2.05	3.7	A-.12	.33	72.5	78.2	Q5
25	23	91	.97	.26	1.38	2.8	1.94	3.7	B-.11	.34	68.1	76.3	Q25
12	28	91	.65	.24	1.11	1.1	1.49	2.6	C .19	.36	70.3	72.3	Q12
6	14	91	1.66	.31	1.12	.7	1.39	1.2	D .13	.29	85.7	84.7	Q6
26	26	91	.78	.25	1.12	1.1	1.16	.9	E .23	.35	68.1	73.8	Q26
1	78	91	-2.37	.32	1.02	.2	1.08	.4	F .25	.30	86.8	85.8	Q1
29	46	91	-.32	.23	1.07	.9	1.05	.5	G .33	.39	60.4	66.6	Q29
21	38	91	.10	.23	.99	-.1	1.06	.5	H .38	.38	70.3	66.8	Q21
16	34	91	.31	.23	1.05	.6	1.02	.2	I .33	.38	67.0	68.5	Q16
3	70	91	-1.71	.27	1.05	.4	1.04	.2	J .30	.35	78.0	78.3	Q3
9	48	91	-.43	.23	1.04	.6	1.02	.2	K .35	.39	64.8	66.9	Q9
10	25	91	.84	.25	1.04	.4	1.04	.3	L .30	.35	73.6	74.6	Q10
22	26	91	.78	.25	1.04	.4	.93	-.3	M .34	.35	70.3	73.8	Q22
2	81	91	-2.70	.35	1.03	.2	.78	-.4	N .28	.27	89.0	89.0	Q2
4	19	91	1.25	.27	1.03	.2	.91	-.3	O .32	.32	76.9	80.1	Q4
24	29	91	.59	.24	1.02	.2	1.00	.0	o .35	.36	67.0	71.6	Q24
30	42	91	-.11	.23	.98	-.2	.93	-.6	n .42	.39	62.6	66.1	Q30
13	23	91	.97	.26	.96	-.3	.91	-.4	m .39	.34	81.3	76.3	Q13
19	50	91	-.53	.23	.94	-.7	.89	-1.0	l .46	.39	72.5	67.5	Q19
17	40	91	-.01	.23	.93	-.9	.92	-.7	k .46	.39	71.4	66.3	Q17
11	36	91	.20	.23	.92	-.9	.88	-1.0	j .46	.38	70.3	67.6	Q11
15	45	91	-.27	.23	.92	-1.0	.86	-1.3	i .48	.39	70.3	66.4	Q15
8	33	91	.37	.24	.90	-1.1	.92	-.5	h .47	.37	72.5	68.9	Q8
14	44	91	-.22	.23	.91	-1.2	.86	-1.3	g .49	.39	67.0	66.3	Q14
28	49	91	-.48	.23	.90	-1.3	.88	-1.1	f .49	.39	75.8	67.2	Q28
7	57	91	-.91	.24	.89	-1.2	.85	-1.1	e .50	.39	75.8	69.8	Q7
23	26	91	.78	.25	.87	-1.1	.88	-.6	d .47	.35	79.1	73.8	Q23
18	47	91	-.37	.23	.86	-1.8	.82	-1.8	c .54	.39	74.7	66.8	Q18
27	43	91	-.17	.23	.82	-2.5	.76	-2.4	b .58	.39	72.5	66.2	Q27
20	54	91	-.74	.23	.77	-2.8	.71	-2.6	a .62	.39	80.2	68.6	Q20
MEAN	39.8	91.0	.00	.25	1.00	-.2	1.03	-.1			73.2	72.2	
S.D.	16.5	.0	.99	.03	.13	1.2	.30	1.4			6.7	6.4	

Figure 2. Item Fit Order.

Based on Figure 2, it can be seen that some questions do not meet the criteria. Analysis of questions that do not meet the criteria is in Table 3.

Table 3. Order Fit Item Interpretation

No.	MNSQ	ZSTD	Pt Measure Corr
5	Not Met	Not Met	Not Met
25	Not Met	Not Met	Not Met
12	Meets	Not Met	Not Met

The item validity test using the Rasch Model involves analyzing the values of Outfit MNSQ, Outfit ZSTD, and Pt Measure Corr. Outfit MNSQ is used to measure how far the participants' responses deviate from the expected model, while Outfit ZSTD provides an indication of the magnitude of that deviation. On the other hand, Pt Measure Corr shows the strength of the relationship between item scores and the overall ability of the participants. Based on the results presented in Table 3, questions number 5, 25, and 12 were found to be inconsistent with the respondents' answers. These three items are categorized as unfit and need to be revised or removed. Questions 5 and 25 will be removed because they do not meet the criteria for all three values tested: Outfit MNSQ, Outfit ZSTD, and Pt Measure

Corr. Meanwhile, question number 12 will be retained on the condition that it is revised, as its Outfit MNSQ value (1.49) is within the acceptable limit (1.5), even though the Outfit ZSTD and Pt Measure Corr values do not meet the desired criteria. In general, an item is considered acceptable if it meets at least one of the three criteria tested (Outfit MNSQ, Outfit ZSTD, and Pt Measure Corr) (Dewi et al., 2021). Therefore, item questions that meet only one criterion, such as Outfit MNSQ, but do not meet the other criteria, should be further analyzed before deciding whether to retain or revise them (Adimayuda et al., 2020).

4.3. Wright Map

The Wright Map illustrates the distribution of item difficulty on the right and student ability on the left. The higher the distribution of item difficulty, the more challenging the question, while items with lower difficulty are considered easier. The results of this analysis can be seen in Figure 3, which shows the relationship between item difficulty and student ability.

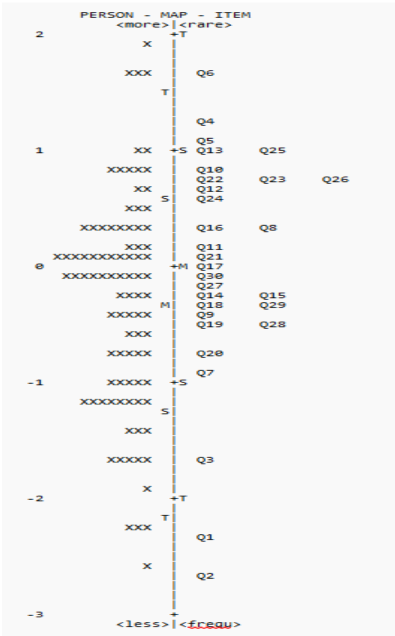

Figure 3. Wright Map

Based on Figure 3, it can be observed that question number 6 is a difficult question. Questions number 2 and 1 are considered easy or very easy. From these results, it is necessary to re-analyze the items with both high and low levels of difficulty.

4.4. Item Measure

Item Measure is used to assess the difficulty level of the question items, as indicated by the logits value. A high logits value signifies the highest level of difficulty for the question, while a lower value indicates an easier question. The criteria for determining the difficulty level of the item questions can be seen in Table 3 (Saely & Shaleh, 2023). The results of the students' answers can be observed in Figure 4.

Table 4. Classification of Question Difficulty Levels

Divisibility Index	Classification
Measure value > 1	Very difficult
0 < Measure value < 1	Difficult
-1 < Measure value < 0	Easy
Measure value < -1	Very easy
Measure value > 1	Very difficult

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
6	14	91	1.66	.31	1.12	.7	1.39	1.2	.13	.29	85.7	84.7	Q6
4	19	91	1.25	.27	1.03	.2	.91	-.3	.32	.32	76.9	80.1	Q4
5	21	91	1.10	.27	1.36	2.4	2.05	3.7	-.12	.33	72.5	78.2	Q5
13	23	91	.97	.26	.96	-.3	.91	-.4	.39	.34	81.3	76.3	Q13
25	23	91	.97	.26	1.38	2.8	1.94	3.7	-.11	.34	68.1	76.3	Q25
10	25	91	.84	.25	1.04	.4	1.04	.3	.30	.35	73.6	74.6	Q10
22	26	91	.78	.25	1.04	.4	.93	-.3	.34	.35	70.3	73.8	Q22
23	26	91	.78	.25	.87	-1.1	.88	-.6	.47	.35	79.1	73.8	Q23
26	26	91	.78	.25	1.12	1.1	1.16	.9	.23	.35	68.1	73.8	Q26
12	28	91	.65	.24	1.11	1.1	1.49	2.6	.19	.36	70.3	72.3	Q12
24	29	91	.59	.24	1.02	.2	1.00	.0	.35	.36	67.0	71.6	Q24
8	33	91	.37	.24	.90	-1.1	.92	-.5	.47	.37	72.5	68.9	Q8
16	34	91	.31	.23	1.05	.6	1.02	.2	.33	.38	67.0	68.5	Q16
11	36	91	.20	.23	.92	-.9	.88	-1.0	.46	.38	70.3	67.6	Q11
21	38	91	.10	.23	.99	-.1	1.06	.5	.38	.38	70.3	66.8	Q21
17	40	91	-.01	.23	.93	-.9	.92	-.7	.46	.39	71.4	66.3	Q17
30	42	91	-.11	.23	.98	-.2	.93	-.6	.42	.39	62.6	66.1	Q30
27	43	91	-.17	.23	.82	-2.5	.76	-2.4	.58	.39	72.5	66.2	Q27
14	44	91	-.22	.23	.91	-1.2	.86	-1.3	.49	.39	67.0	66.3	Q14
15	45	91	-.27	.23	.92	-1.0	.86	-1.3	.48	.39	70.3	66.4	Q15
29	46	91	-.32	.23	1.07	.9	1.05	.5	.33	.39	60.4	66.6	Q29
18	47	91	-.37	.23	.86	-1.8	.82	-1.8	.54	.39	74.7	66.8	Q18
9	48	91	-.43	.23	1.04	.6	1.02	.2	.35	.39	64.8	66.9	Q9
28	49	91	-.48	.23	.90	-1.3	.88	-1.1	.49	.39	75.8	67.2	Q28
19	50	91	-.53	.23	.94	-.7	.89	-1.0	.46	.39	72.5	67.5	Q19
20	54	91	-.74	.23	.77	-2.8	.71	-2.6	.62	.39	80.2	68.6	Q20
7	57	91	-.91	.24	.89	-1.2	.85	-1.1	.50	.39	75.8	69.8	Q7
3	70	91	-1.71	.27	1.05	.4	1.04	.2	.30	.35	78.0	78.3	Q3
1	78	91	-2.37	.32	1.02	.2	1.08	.4	.25	.30	86.8	85.8	Q1
2	81	91	-2.70	.35	1.03	.2	.78	-.4	.28	.27	89.0	89.0	Q2
MEAN	39.8	91.0	.00	.25	1.00	-.2	1.03	-.1			73.2	72.2	
S.D.	16.5	.0	.99	.03	.13	1.2	.30	1.4			6.7	6.4	

Figure 4. Question Difficulty Level

The Item Measure results in Figure 3.4 show that questions with very difficult categories are at a Measure value> 1, namely in question numbers 6, 4, and 5. Questions with difficult categories are at a value of 0 < Measure value < 1, namely in question numbers 13, 25, 10, 22, 23, 26, 12, 24, 8, 16, 11, and 21. Questions in the easy category are at a value of -1 < Measure value < 0, namely in question numbers 17, 30, 27, 14, 15, 29, 18, 9, 28, 19, 20, and 7. Questions with a very easy level, namely Measure value < -1, namely in numbers 3, 1, and 2.

The relationship between test items and student abilities can be analyzed through the Wright Map, as shown in Figure 3. This map illustrates the distribution of item difficulty levels, which corresponds to the item measure values displayed in Figure 4. These values categorize questions from very easy to very difficult. The analysis reveals that item number 6 falls into the very difficult category, exceeding the standard deviation threshold (T), while items 1 and 2 are classified as very easy. Despite the high difficulty of item 6, some students demonstrated abilities above its difficulty level. Out of 91 students, only 14 answered item 6 correctly, confirming its high difficulty level. The classification of item difficulty is determined based on the item measure values presented in logit units, as outlined in Table 4. Since logit values operate on the same scale, they allow for direct comparison of difficulty across items. In the Rasch Model, the difficulty of each item is evaluated based on its measure value in logits, where a higher measure indicates a higher level of ability required to answer the item correctly, and vice versa (S. N. Azizah et al., 2022)

4.5. Undimentionality Analysis

Unidimensionality analysis aims to ensure that the instrument consistently measures a single underlying construct. This assessment is based on the value of the raw variance explained by measures and the unexplained variance in the first to fifth contrasts. An instrument is considered to meet unidimensionality criteria if the raw variance explained by measures is at least 20%. The classification of this value is as follows: "Fair" for values between 20%–40%, "Good" for 40%–60%, and "Excellent" if it exceeds 60%. Furthermore, to support unidimensionality, the unexplained variance in each of the first through fifth contrasts should be less than 15%. These criteria help determine the extent to which residual factors may affect measurement results, as explained by Fulmer (2016).

```
 CONTRAST 5 FROM PRINCIPAL COMPONENT ANALYSIS
     Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                                -- Empirical --    Modeled
Total raw variance in observations     =        40.6 100.0%         100.0%
  Raw variance explained by measures   =        10.6  26.2%          25.5%
    Raw variance explained by persons  =         3.7   9.1%           8.9%
    Raw Variance explained by items    =         6.9  17.1%          16.7%
  Raw unexplained variance (total)     =        30.0  73.8% 100.0%   74.5%
    Unexplned variance in 1st contrast =         2.9   7.0%   9.5%
    Unexplned variance in 2nd contrast =         2.5   6.2%   8.4%
    Unexplned variance in 3rd contrast =         2.2   5.5%   7.5%
    Unexplned variance in 4th contrast =         1.8   4.5%   6.2%
    Unexplned variance in 5th contrast =         1.7   4.3%   5.8%
```

Figure 5. Undimensionality analysis

Based on the results presented in Figure 5, the raw variance explained by measures is 26.2%, which falls into the "Good" category. This indicates that the instrument adequately explains a single underlying construct. Meanwhile, the unexplained variance in the 1st through 5th contrasts of the residuals are as follows: 7.0% in the first contrast, 6.2% in the second, 5.5% in the third, 4.5% in the fourth, and 4.3% in the fifth. All residual values are below the 15% threshold, suggesting that the instrument does not contain significant latent dimensions beyond the primary one. Therefore, the instrument meets the unidimensionality criteria in accordance with the standards of Rasch Model analysis.

4.6. Information Function Test

Figures 6 and 7 present the information function of the test instrument based on the Rasch Model analysis. On these graphs, the x-axis denotes the spectrum of students' abilities, while the y-axis indicates the amount of information the test provides at each ability level. The data reveal that the test offers high information across both lower and higher levels of student ability, suggesting that the instrument performs effectively across a broad range of learner competencies (Azizah & Wahyuningsih, 2020).

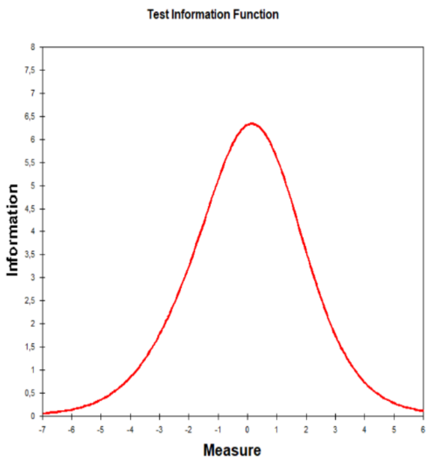

Figure 6. Information Function Tes

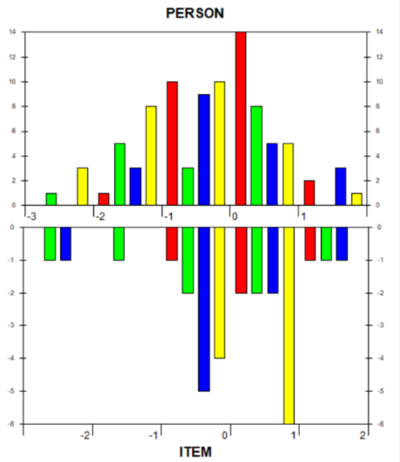

Figure 7. Person-Item Barchart

These findings highlight the instrument's potential for reliably measuring students' conceptual understanding of dynamic electricity. The strong measurement precision across varying ability levels suggests that the instrument is both appropriate and effective. Consequently, it can be further refined and utilized in future assessments to support deeper diagnostic evaluation in physics education contexts.

5. Conclusions

Based on the analysis conducted, the web-based 4-Tier instrument on dynamic electricity material meets the established criteria for validity. The items within this instrument are classified as fit, in accordance with the set standards. With a Cronbach Alpha value of 0.78, indicating the quality of the interaction between participants and items, this instrument is categorized as "Good." Meanwhile, the Person Reliability value ranging from 0.76 to 0.78 falls within the "Fair" category, and the Item Reliability values of 0.93 to 0.94 are categorized as "Excellent." However, three question items require reassessment as they do not meet the expected criteria, and therefore need to be revised or deleted. Overall, the web-based 4-Tier instrument on dynamic electricity material is considered a valid, reliable, and effective tool for measuring students' understanding.

References

[1] Adimayuda, R., Aminudin, A. H., Kaniawati, I., Suhendi, E., & Samsudin, A., "A multitier open ended momentum and impulse (MOMI) instrument: Developing and assessing quality of conception of 11th grade Sundanese students with Rasch analysis," *International Journal of Scientific and Technology Research*, vol. 9, no. 2, pp. 4799–4804, 2020. doi: 10.17478/jegys.574524.

[2] Anggraini, N., Iswanto, B. H., & Wibowo, F. C., "Four Tier Test (FTT) development in the form of virtualization static fluid test (VSFT) using Rasch model analysis to support learning during the Covid-19 pandemic," *Journal of Physics: Conference Series*, vol. 2019, no. 1, 2021. doi: 10.1088/1742-6596/2019/1/012005.

[3] Anggrayni, S., & Ermawati, F. U., "The validity of Four-Tier's misconception diagnostic test for work and energy concepts," *Journal of Physics: Conference Series*, vol. 1171, no. 1, 2019. doi: 10.1088/17426596/1171/1/012037.

[4] Arcelay, I., Goti, A., Oyarbide-Zubillaga, A., Akyazi, T., Alberdi, E., & Garcia-Bringas, P., "Definition of the future skills needs of job profiles in the renewable energy sector," *Energies*, vol. 14, no. 9, p. 2609, 2021. doi: 10.3390/en14092609.

[5] Astuti, A. P., Aziz, A., Sumarti, S. S., & Bharati, D. A. L., "Preparing 21st Century Teachers: Implementation of 4C Character's Pre-Service Teacher through Teaching Practice," *Journal of Physics: Conference Series*, vol. 1233, no. 1, 2019. doi: 10.1088/1742-6596/1233/1/012109.

[6] Azizah, A., & Wahyuningsih, S., "Penggunaan Model Rasch Untuk Analisis Instrumen Tes Pada Mata Kuliah Matematika Aktuaria," *JUPITEK: Jurnal Pendidikan Matematika*, vol. 3, no. 1, pp. 45–50, 2020. doi: 10.30598/jupitekvol3iss1pp45-50.

[7] Azizah, A., & Wahyuningsih, S., "Use of Rasch Model for Test Instrument Analysis in Actuarial Mathematics Course," *JUPITEK: Jurnal Pendidikan Matematika*, vol. 3, no. 1, pp. 45–50, 2020.

[8] Azizah, S. N., Akhsan, H., Muslim, M., & Ariska, M., "Analysis of college students misconceptions in astronomy using four-tier test," *Journal of Physics: Conference Series*, vol. 2165, no. 1, 2022. doi: 10.1088/1742-6596/2165/1/012004.

[9] Bodzin, A., Hammond, T., Fu, Q., & Farina, W., "Development of Instruments to Assess Students' Spatial Learning Attitudes (SLA) and Interest in Science, Technology and Geospatial Technology (STEM-GEO)," *International Journal of Educational Methodology*, vol. 6, no. 1, pp. 67–81, 2020. doi: 10.12973/ijem.6.1.67.

[10] Dewi, F. H., Samsudin, A., & Chandra, D. T., "Developing FD-MT to investigate students' mental model on fluid dynamic concept: a Rasch model analysis," *Journal of Physics: Conference Series*, vol. 2098, no. 1, 2021. doi: 10.1088/1742-6596/2098/1/012020.

[11] Diani, R., Alfin, J., Anggraeni, Y. M., Mustari, M., & Fujiani, D. Mira Amelia, Mary Margaret Thomas, Lilit Rusyati, "Four-tier diagnostic test with certainty of response index on the concepts of fluid," *Journal of Physics: Conference Series*, vol. 1155, no. 1, 2019. doi: 10.1088/1742-6596/1155/1/012078.

[12] Giancoli, D. C., *Physics: Principles with Applications*, 7th ed. United States: Pearson Education, Inc., 2016.

[13] Istiyono, E., Dwandaru, W. S. B., Ayub, M. R. S., Saepuzaman, D., Zakwandi, R., Rachman, A., Yusron, E., & Santoso, P. H., "A study of the readiness of post-pandemic computer-based four-tier diagnostic test (CBFTDT): A review of economic level, school grades, and device accessibility," *Jurnal Penelitian Dan Evaluasi Pendidikan*, vol. 27, no. 1, pp. 92–104, 2023. doi: 10.21831/pep.v27i1.56005.

[14] Kaniawati, I., Rahmadani, S., Fratiwi, N. J., Suyana, I., Danawan, A., Samsudin, A., & Suhendi, E., "An analysis misconceptions of about students' the implementation of active learning of optics and photonics approach assisted by computer simulation," *International Journal of Emerging Technologies in Learning*, vol. 15, no. 9, pp. 76–93, 2020. doi: 10.3991/ijet.v15i09.

[15] Koçak, D., "Investigation of rater tendencies and reliability in different assessment methods with many facet Rasch model," *International Electronic Journal of Elementary Education*, vol. 12, no. 4, pp. 349–358, 2020. doi: 10.26822/iejee.2020459464.

[16] Lengkong, M., Istiyono, E., Rampean, B. A. O., Tumanggor, A. M. R., & Nirmala, M. F. T., "Development of Two-Tier Test Instruments to Detect Student's Physics Misconception," *Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)*, vol. 528, pp. 561–566, 2021. doi: 10.2991/assehr.k.210305.082.

[17] Maison, Safitri, I. C., & Rendy, W. W., "Temperature and calor topic using four-tier diagnostic," *Jurnal Ilmiah Pendidikan Fisika*, vol. 11, no. 2, pp. 195–202, 2019.

[18] Martin, F., Stamper, B., & Flowers, C., "Examining student perception of readiness for online learning: Importance and confidence," *Online Learning*, vol. 24, no. 2, pp. 38–58, 2020. doi: 10.24059/olj.v24i2.2053.

[19] Mufit, F., Festiyed, Fauzan, A., & Lufri, "The application of real experiments video analysis in the CCBL model to remediate the misconceptions about motion's concept," *Journal of Physics: Conference Series*, vol. 1317, no. 1, 2019. doi: 10.1088/1742-6596/1317/1/012156.

[20] Soeharto, S., "Development of A Diagnostic Assessment Test to Evaluate Science Misconceptions in Terms of School Grades: A Rasch Measurement Approach," *Journal of Turkish Science Education*, vol. 18, no. 3, pp. 351–370, 2021. doi: 10.36681/tused.2021.78.

[21] Suryana, T. G. S., Setyadin, A. H., Samsudin, A., & Kaniawati, I., "Assessing Multidimensional Energy Literacy of High School Students: An Analysis of Rasch Model," *Journal of Physics: Conference Series*, vol. 1467, no. 1, 2020. doi: 10.1088/1742-6596/1467/1/012034.

[22] Tumanggor, A. M. R., Supahar, S., Ringo, E. S., & Harliadi, M. D., "Detecting Students' Misconception in Simple Harmonic Motion Concepts Using Four-Tier Diagnostic Test Instruments," *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, vol. 9, no. 1, pp. 21–31, 2020. doi: 10.24042/jipfalbiruni.v9i1.4571.