Journal of Innovative Physics Education Research (JIPER)
e-ISSN: 3089-5006 *Vol. 1 No. 1, April 2025, Page 28-38*

Development and Limited Test of Kinematics Conception Diagnostic Instrument: A Rasch Model Analysis Approach

Hana Apriliana*, Aden Yusril Sasmita, Ilham Mustaqim, Rika Rahmawati, Tresna Galih Sukma Suryana
Department of Physics Education, Universitas Sultan Ageng Tirtayasa
Corresponding Author e-mail: *2280220039@untirta.ac.id

Received January 10, 2025; Revised January 30, 2025; Accepted March 26, 2025; Available online April 16, 2025
DOI: https://doi.org/10.61142/jiper.v1i1.194

Keywords :
Kinematics, Misconception, Students

ABSTRACT
This study aims to identify and analyze misconceptions experienced by students on kinematics material. The approach used is quantitative, with a two-level diagnostic test method consisting of 10 multiple choice questions designed to detect misconceptions, involving 30 students of class XI MIPA at SMAN 4 Serang City. The research process included instrument development, expert validity test, limited trial, and data analysis. The question testing design used the Rasch Model Analysis method. Overall, this diagnostic test instrument showed good validity. The analysis results showed that the item reliability value reached 0.89, which indicated that the quality of the items was good. The findings of the data analysis show that students experience misconceptions caused by misunderstanding the concept of kinematics. Based on the results of this study, it is recommended that more effective learning strategies, such as the use of visual media and hands-on practicum, be applied to improve students' understanding of the concept of kinematics. In addition, training for teachers is also needed to improve their ability to identify and overcome misconceptions experienced by students.

INTRODUCTION

Physics is a branch of natural science that has an important role in explaining various phenomena in nature through concept-based and quantitative approaches. Understanding the basic concepts in physics is very important, considering that physics is not only theoretical, but also an applied science that is closely related to daily events and the surrounding environment [13]. Therefore, students' ability to understand, identify, and interpret physics concepts is very important so that they are able to make logical reasoning and solve physics problems systematically [10]. However, the reality in the field shows that not all students are able to master physics concepts well. Often found students who have a strong belief in a concept that is actually not in accordance with scientific concepts. This condition is known as misconception [9], which is an understanding that deviates from the scientific truth that has been generally accepted by experts [18]. These misconceptions can stem from various things, such as inappropriate learning experiences, errors in interpreting natural phenomena, or ineffective teaching approaches.

One of the most vulnerable physics materials to cause misconceptions is kinematics. Kinematics itself is a study of the motion of objects without considering the cause of the motion, and is an important basis for understanding other more complex physics concepts. A misunderstanding of this material can have an impact on students' inability to master advanced material. Factors that trigger the emergence of misconceptions in this topic include limited media and learning resources, inappropriate learning methods, and the unavailability of accurate evaluation tools to detect misconceptions. To overcome these problems, an evaluation instrument is needed that not only measures learning outcomes, but is also able to detect and analyze students' misconceptions. One method that is considered effective is the use of concept understanding-based diagnostic tests with a quantitative approach. This test is more efficient than other approaches such as interviews or concept maps because it is able to identify misconceptions thoroughly in a relatively short time [22]. This instrument is designed to measure students' understanding in depth and provide a clear picture of the location of their conceptual errors.

This study aims to develop and conduct a limited trial of a concept understanding diagnostic test instrument on kinematics material. It is hoped that the existence of this instrument can help teachers identify misconceptions that occur in students and develop more effective learning strategies. In addition, this instrument can also be utilized by students as a reflection tool to assess and improve their understanding independently. The results of this study are expected to make a real contribution to improving the quality of kinematics learning at the secondary school level, as well as being the basis for developing diagnostic instruments in other physics materials and science in general.

METHOD

A. Research Design

This study adopts the Research and Development (R&D) model with a quantitative method with a descriptive approach that aims to evaluate a two-tier diagnostic test instrument on the topic of kinematics. This approach was chosen because it allows for objective and measurable data analysis in the form of numbers. To analyze the data, the Rasch Model was used to evaluate the quality of each item based on indicators of item validity, reliability, difficulty level, differentiating power, and the effectiveness of distractors on the instrument developed.

B. Participants

The subjects in this study were 30 students of class XI SMAN 4 Serang City who had obtained learning on kinematics material. The sampling technique used was purposive sampling, which is the selection of samples intentionally based on certain considerations. The criteria used include location affordability, subject willingness to participate, and representation of population characteristics relevant to the research objectives.

C. Data Collection Procedure

Data collection procedures include:

1. Instrument Development
    - Determining indicators and aspects of conception based on the curriculum and literature review.
    - Developing two-tier items based on the indicators that have been determined.
    - Content validation by experts (lecturers of evaluation courses).
2. Limited Trial
    - The instrument was tested on a number of students to collect initial empirical data.
    - Data from the pilot test was used to evaluate the instrument using the Rasch Model.
3. Data Analysis

Data were analyzed using Rasch Model support software with WINSTEPS.

D. Research Instrument

The instrument used in this study is the Two-Tier Diagnostic Test, which consists of two levels :

- The first level contains 10 multiple choice questions to test students' concept understanding.
- The second level contains reasoning questions that refer to the answers in the first level.

E. Data Analysis

Data analysis was conducted to determine :

a) Validity Test

The criteria used to check the suitability of outlying items (outliers or misfits) are:

1. The corresponding Outfit Mean Square (MNSQ) value: 0.5 < MNSQ < 1.5.
2. Corresponding Z-standard Outfit Value (ZSTD): -2.0 < ZSTD < 2.0.

If the item does not meet these three criteria, then the item is said to be poor and needs to be improved or replaced. This will ensure in the future that students' level of understanding is indeed tested through appropriate and quality items.

Table 1. Validity test interpretation

Validity	interpretation
0,00 – 0,20	Very Low (SR)
0,20 – 0,40	Low (R)
0,41 – 0,60	Fair (C)
0,61 – 0,80	High (T)
0,81 – 1,00	Very high (ST)

Validity formula :

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}} \quad (1)$$

b) Reliability Test

Fundamental Statistics in Psychology and Education says that very low reliability with a coefficient value <0.50 indicates very low internal consistency [6]. The coefficient value between 0.50 - 0.59 is considered to have low reliability, indicating insufficient consistency. The coefficient between 0.60 - 0.69 is considered moderate, indicating a fairly significant consistency. A coefficient value between 0.70 - 0.79 is considered high, indicating strong consistency. Meanwhile, a coefficient value of => 0.80 or higher indicates very high reliability and very strong consistency.

Table 2. Interpretation of person reliability and item reliability values

Reliability	Interpretation
0,94 < Value	Special
0,90 – 0,94	Very Good
0,80 – 0,90	Good
0,67 – 0,80	Fair
Value < 0,67	Weak

K-R20 formula :

$$r_{11} = \left(\frac{k}{k-1}\right)\left(\frac{SB^2 - \sum pq}{SB^2}\right) \qquad (2)$$

Description :

r_{11} = overall test reliability
p = proportion of subjects who answered the item correctly
q = proportion of subjects who answered the item incorrectly (q=1-p)
$\sum pq$ = the sum of the product of p and q
n = number of items
S = Standard Deviation of the Test (standard deviation is the alar of variance)

c) Test the level of difficulty

Item difficulty levels are grouped into three main categories, namely easy, medium, and difficult. Questions are said to be easy if more than 70% of test takers answer correctly, moderate if between 30%-70% of participants answer correctly, and difficult if less than 30% of participants answer correctly.

Table 3. Interpretation of difficulty level

Difficulty Level (TK)	Interpretation
0,86 < TK	Very difficult
0,00 – 0,86	Difficult
-0,86 – 0,00	Easy
TK < -0,86	Very easy

Formula Level of difficulty :

$$TK = \frac{Number\ of\ students\ answering\ correctly}{Number\ of\ students\ taking\ the\ test} \qquad (3)$$

d) Distinguishing Power Test

To be acceptable, the D-value (discrimination) is 0.30 or above. While to be considered satisfactory is 0.40 and above.

Table 4. Interpretation of distinguishing power

PT MEASURE-AL CORR (ID)	Interpretation
0,40 < ID	Very good
0,30 – 0,40	Good
0,20 – 0,30	Not so good
ID < 0,20	Poor

Distinguishing Power Formula:

$$DP = \frac{(H - L)}{K} \tag{4}$$

Description :
H = Number of Upper Group who answered correctly
L = Number of Lower Group who answered correctly
K = Number of students in each group

e) Distractor test

A distractor is said to be functioning properly if it is selected by at least 5% of the students taking the test. Distractor analysis is essential in the development of high-quality items as it can help ensure that each option can contribute to fair and accurate scoring.

RESULTS AND DISCUSSIONS

The results of data processing analysis in the form of image-shaped output obtained from Winstep Rasch 4.4.3 software which can be used to see the validity of the instrument, instrument reliability, instrument difficulty level, instrument distinguishing power, and instrument distractors.

a) Instrument Validity Test

Validity is an indicator of the extent to which an instrument is able to measure in accordance with its measurement objectives [4]. Instruments that have good validity will produce accurate data and represent the actual conditions. Referring to Psychological Testing, the level of validity can be determined from the correlation value, where values between 0.60 to 0.79 reflect high validity with a strong relationship between variables [3]. Whereas correlation values ≥ 0.80 indicate very high validity with a very strong relationship. To ensure that the instrument truly measures aspects of student resilience, validity testing was conducted. In the Winsteps application, Table 10 (Item Fit Order) was used to analyze item validity based on Outfit Mean Square (MNSQ) and Outfit Z-Standard (ZSTD) values. Through this analysis, items can be classified as valid (fit) or invalid (misfit). Items that do not meet the criteria need to be revised or replaced to maintain the quality and accuracy of the instrument.The following is an image of the output from table 10 which shows information related to the validity criteria of the instrument used:

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%	Item
9	14	27	-.07	.42	1.25	1.80	1.65	2.53	A .07	.38	63.0	65.7	S9
2	26	30	-2.15	.57	1.23	.69	1.12	.39	B .11	.30	83.3	87.2	S2
7	16	29	-.20	.40	1.00	.04	.97	-.07	C .36	.36	62.1	65.3	S7
8	12	27	.32	.42	.89	-.79	.82	-.75	D .48	.37	74.1	64.4	S8
10	8	28	1.05	.45	.89	-.47	.88	-.24	c .46	.37	85.7	75.0	S10
5	21	25	-1.90	.58	.88	-.20	.66	-.41	b .45	.32	88.0	85.1	S5
3	2	29	2.95	.76	.68	-.41	.28	-.68	a .56	.25	93.1	93.0	S3
MEAN	12.9	28.1	.49	.91	.98	.10	.91	.11			78.5	76.5	
P.SD	10.1	1.4	3.01	.61	.19	.82	.39	1.05			11.4	11.0	

Fig 1. *Item fit order output* in *Winstep*

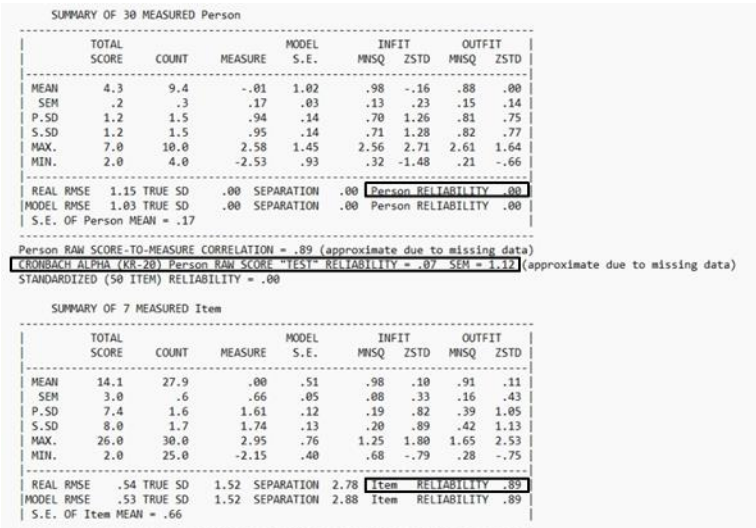
From the results of the data analysis above, in OUTFIT MNSQ there is 1 item that does not meet the criteria, namely question number 9 with an MNSQ value of 1.65, but overall the resulting MNSQ value is 1.07, which means it is accepted. In OUTFIT ZSTD there is 1 item that does not meet the criteria, namely question number 9 with an MNSQ value of 2.53, but it can still be said to be valid because overall it is still within the accepted value range of 0.11.

b) Instrument Reliability Test

Reliability refers to the extent to which a measuring instrument shows consistency when used repeatedly [16]. A test is said to be reliable if it is able to produce scores that are relatively stable, even though it is carried out in different situations or times. In the Rasch approach, reliability analysis includes item reliability, respondent reliability, and Cronbach's Alpha, which describes the overall relationship between test takers and question items.

According to Fundamental Statistics in Psychology and Education, reliability is divided into several categories. A coefficient below 0.50 indicates a very low level of internal consistency [6]. A coefficient between 0.50 to 0.59 indicates low reliability. Meanwhile, a coefficient of 0.60-0.69 indicates moderate reliability, 0.70-0.79 indicates high reliability, and a coefficient above 0.80 reflects very high reliability with strong consistency.

Reliability testing is conducted to determine the extent to which the instrument can be trusted and consistent as a measuring tool. In the Winsteps application, this analysis is carried out using Table 3.1 (Summary Statistic), which displays the reliability value both in terms of test participants (person reliability) and items (item reliability).The following is an image of the summary statistic output that shows information about the reliability of the instrument used:



```
     SUMMARY OF 30 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN       4.3       9.4        -.01    1.02       .98   -.16    .88    .00 |
|  SEM        .2        .3         .17     .03       .13    .23    .15    .14 |
| P.SD       1.2       1.5         .94     .14       .70   1.26    .81    .75 |
| S.SD       1.2       1.5         .95     .14       .71   1.28    .82    .77 |
| MAX.       7.0      10.0        2.58    1.45      2.56   2.71   2.61   1.64 |
| MIN.       2.0       4.0       -2.53     .93       .32  -1.48    .21   -.66 |
|-----------------------------------------------------------------------------|
| REAL RMSE    1.15 TRUE SD     .00  SEPARATION    .00  Person RELIABILITY  .00 |
|MODEL RMSE    1.03 TRUE SD     .00  SEPARATION    .00  Person RELIABILITY  .00 |
| S.E. OF Person MEAN = .17                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .89 (approximate due to missing data)
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .07  SEM = 1.12 (approximate due to missing data)
STANDARDIZED (50 ITEM) RELIABILITY = .00

     SUMMARY OF 7 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      14.1      27.9         .00     .51       .98    .10    .91    .11 |
|  SEM       3.0        .6         .66     .05       .08    .33    .16    .43 |
| P.SD       7.4       1.6        1.61     .12       .19    .82    .39   1.05 |
| S.SD       8.0       1.7        1.74     .13       .20    .89    .42   1.13 |
| MAX.      26.0      30.0        2.95     .76      1.25   1.80   1.65   2.53 |
| MIN.       2.0      25.0       -2.15     .40       .68   -.79    .28   -.75 |
|-----------------------------------------------------------------------------|
| REAL RMSE     .54 TRUE SD    1.52  SEPARATION   2.78  Item   RELIABILITY  .89 |
|MODEL RMSE     .53 TRUE SD    1.52  SEPARATION   2.88  Item   RELIABILITY  .89 |
| S.E. OF Item MEAN = .66                                                     |
-------------------------------------------------------------------------------
```

Fig 2. Summary statistic output in Winsteps

From the results of the data analysis above, the Person Reability test results have a value of 0.00, which means that the consistency of student answers to questions is in the weak category. The Item Reability test results have a value of 0.89, which means that the quality of the items in the instrument is in the good category. Meanwhile, the Cronbach Alpha value obtained is 0.07 which is in the very low (poor) category.

c) Instrument Difficulty Test

The difficulty level of the item is the proportion between the number of test takers who answer the item correctly and the number of test takers [14]. This means that the more test takers who answer the item correctly, the greater the difficulty index, which means the easier the question is. Conversely, fewer test takers who answer the item correctly, the more difficult the item. The test instrument difficulty level is the proportion between the number of test takers who answered the instrument correctly and the number of test takers. This means that the more test takers who answer the items correctly, the greater the difficulty index, which means the easier the question is. Conversely, the fewer test participants who answer the items correctly, the more difficult the question is.

The difficulty level of the instrument is grouped into three main categories, namely easy, medium, and difficult. The instrument is said to be easy if more than 70% of test takers answer correctly, moderate if between 30%-70% of participants answer correctly, and difficult if less than 30% of participants answer correctly. The level of difficulty test used table 13 (Fig. 3) in the Winsteps program.

Item STATISTICS: MEASURE ORDER

ENTRY	TOTAL	TOTAL	JMLE	MODEL	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Item
4	0	28	4.97	1.83	MAXIMUM MEASURE				.00	.00	100.0	100.0	S4
6	0	28	4.97	1.83	MAXIMUM MEASURE				.00	.00	100.0	100.0	S6
3	2	29	2.95	.76	.68	-.41	.28	-.68	.56	.25	93.1	93.0	S3
10	8	28	1.05	.45	.89	-.47	.88	-.24	.46	.37	85.7	75.0	S10
8	12	27	.32	.42	.89	-.79	.82	-.75	.48	.37	74.1	64.4	S8
9	14	27	-.07	.42	1.25	1.80	1.65	2.53	.07	.38	63.0	65.7	S9
7	16	29	-.20	.40	1.00	.04	.97	-.07	.36	.36	62.1	65.3	S7
5	21	25	-1.90	.58	.88	-.20	.66	-.41	.45	.32	88.0	85.1	S5
2	26	30	-2.15	.57	1.23	.69	1.12	.39	.11	.30	83.3	87.2	S2
1	30	30	-5.03	1.83	MINIMUM MEASURE				.00	.00	100.0	100.0	S1
MEAN	12.9	28.1	.49	.91	.98	.10	.91	.11			78.5	76.5	
P.SD	10.1	1.4	3.01	.61	.19	.82	.39	1.05			11.4	11.0	

Fig 3. JMLE MEASURE output in Winsteps

From the results of the data analysis above, in JMLE MEASURE there are 4 items in the very difficult category (question numbers 4, 3, 6 and 10), 1 item in the difficult category (question number 8), 2 items in the easy category (question numbers 7 and 9) and 3 items in the very easy category (question numbers 1, 2 and 5).

d) Distinguishing power test

The differentiating power of the question is the ability of the question with its score to distinguish test takers from high and low groups. In other words, the higher the discriminating power of the question, the more participants from the high group can answer the question correctly and the fewer test takers from the low group can answer the question correctly [5]. To be acceptable, the value of D (discrimination: question discriminating power) is 0.30 or more. To be considered satisfactory, it should be 0.40 and above.

The differentiating power test is the ability of the question with its score to distinguish test takers from high and low groups. In other words, the higher the discriminating power of the question, the more participants from the high group who can answer the question correctly and the fewer test takers from the low group who can answer the question correctly. To be acceptable, the D value (discrimination: question discriminating power) is 0.30 or more. While to be satisfactory is 0.40 and above. The discriminating power test uses table 3.1 of PT MEASURE-AL CORR (Fig. 4) in the Winsteps program.

ENTRY	TOTAL	TOTAL	JMLE	MODEL	INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		
NUMBER	SCORE	COUNT	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%	Item
9	14	27	-.07	.42	1.25	1.80	1.65	2.53	A .07	.38	63.0	65.7	S9
2	26	30	-2.15	.57	1.23	.69	1.12	.39	B .11	.30	83.3	87.2	S2
7	16	29	-.20	.40	1.00	.04	.97	-.07	C .36	.36	62.1	65.3	S7
8	12	27	.32	.42	.89	-.79	.82	-.75	D .48	.37	74.1	64.4	S8
10	8	28	1.05	.45	.89	-.47	.88	-.24	c .46	.37	85.7	75.0	S10
5	21	25	-1.90	.58	.88	-.20	.66	-.41	b .45	.32	88.0	85.1	S5
3	2	29	2.95	.76	.68	-.41	.28	-.68	a .56	.25	93.1	93.0	S3
MEAN	12.9	28.1	.49	.91	.98	.10	.91	.11			78.5	76.5	
P.SD	10.1	1.4	3.01	.61	.19	.82	.39	1.05			11.4	11.0	

Fig. 4 PT MEASURE-AL CORR output on Winsteps

From the results of the data analysis above, the differentiating power can be seen through PT MEASURE-AL CORR, where there are 4 questions in the excellent category (questions number 3, 5, 8 and 10) and 1 differentiating power in the good category (question number 7), this shows that the question is able to distinguish between students who know the answer correctly and students who cannot answer the question or students who answer incorrectly. However, there were 2 questions in the poor category (questions number 2 and 9), which means that the question is not able to distinguish between students who know the answer correctly and students who cannot answer the question or students who answer incorrectly.

e) Distractor Test

The distractor test aims to assess how well the distractor functions in diverting participants' attention from the correct answer. An effective distractor should be able to attract a certain number of participants who do not understand the material correctly, so that not all participants with insufficient knowledge choose the correct answer by chance. Research on the test of distractors in multiple-choice questions shows that a good distractor analysis can help distinguish between participants who understand and do not understand the material being tested [19].

A distractor is said to be functioning properly if it is selected by at least 5% of the students taking the test. Distractor analysis is essential in the development of high-quality items as it can help ensure that each option can contribute to fair and accurate scoring.

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: MISFIT ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	Item
9 A	E	0	1	4	-1.38	.00		.3	.3	-.28	S9
	C	0	4	15	-1.17	.87	.51	.5	.4	-.52	
	B	0	8	30	.68	1.05	.40	2.3	3.4	.44	
	A	1	14	52	.09*	.38	.10	1.0	1.0	.07	
	MISSING	***	3	10#	-.25	.62	.44			-.09	
2 B	B	0	4	13	-.26	.36	.21	1.5	1.1	-.11	S2
	A	1	26	87	.03	.99	.20	1.4	1.2	.11	
7 C	E	0	9	31	-.74	.79	.28	.7	.6	-.57	S7
	D	0	3	10	.37	.00	.00	1.5	1.4	.12	
	B	0	1	3	1.30	.00		3.8	3.6	.26	
	A	1	16	55	.34*	.79	.20	1.0	.9	.36	
	MISSING	***	1	3#	-1.35	.00				-.27	
8 D	E	0	1	4	-2.53	.00		.1	.1	-.53	S8
	C	0	1	4	-1.38	.00		.2	.2	-.30	
	B	0	6	22	-.58	.24	.11	.5	.5	-.36	
	D	0	7	26	.30	.09	.04	1.2	1.2	.15	
	A	1	12	44	.58	.87	.26	.9	.8	.48	
	MISSING	***	3	10#	-.61	.53	.37			-.21	
10 c	B	0	8	29	-.41	.98	.37	.8	.8	-.26	S10
	A	0	11	39	-.28	.59	.19	.7	.8	-.21	
	C	0	1	4	.37	.00		1.1	1.3	.08	
	D	1	8	29	.68	1.05	.40	.9	.9	.46	
	MISSING	***	2	7#	.18	.05	.05			.05	
5 b	D	0	1	4	-2.53	.00		.1	.1	-.53	S5
	C	0	3	12	-.47	.00	.00	1.2	.8	-.19	
	B	1	21	84	.23	.88	.20	.9	.9	.45	
	MISSING	***	5	17#	-.21	.61	.31			-.10	
3 a	E	0	16	55	-.44	.52	.13	.3	.5	-.49	S3
	A	0	5	17	-.27	1.13	.56	.5	.8	-.12	
	B	0	3	10	.37	.00	.00	.5	1.0	.14	
	C	0	3	10	.99	.44	.31	1.1	2.0	.36	
	D	1	2	7	1.94	.64	.64	.4	.3	.56	
	MISSING	***	1	3#	.23	.00				.05	

Fig 5. MISFIT ORDER output on Winsteps

From the results of the data analysis above, the distractor is said to function properly if it is selected by at least 5% of the students who take the test. In the DATA COUNT percentage, you can see the dysfunctionality of each question.

Table 5. Distractor Analysis Table of Each Item using Ministep Rasch

No. About	Distractor Analysis	Recommendation
2	All distractors did not work, as 87% of students chose the correct answer key and the other 13% chose option B (distractor).	All options except the answer key and option B are replaced.
3	All options on the question (distractors) function.	-
5	Only option C met the functionality of the distractor, at 12%. The other options were below 5% of the voters	Options A, D and E are replaced
7	Only options D and E met the functionality of the distractors, with 10% and 31% voters respectively. The other options were below 5% voter turnout	Option B dan C replaced
8	Only options B and D met the functionality of the distractors, with 22% and 26% of voters respectively. Options C and E were selected by less than 5%	Option C and E replaced
9	Only options B and C met the functionality of the distractors, with 30% and 15% voters respectively. The other options were below 5% of voters	Option D and E replaced
10	Only options A and B met the functionality of the distractors, at 39% and 29%. Options C and E were selected by less than 5%	Option C and E replaced

Through the results of this test can be an important reflection material for teachers in improving the quality of the learning process. If it is known that many students have difficulty on questions with moderate difficulty, then this can be an indicator of an understanding gap that needs to be addressed immediately. Thus, the use of the Rasch Model not only plays a role in assessment evaluation, but also as a strategic tool in improving the effectiveness and quality of physics learning as a whole.

CONCLUSION AND SUGGESTION

The limited trial conducted shows that the two-tier diagnostic test instrument developed has met good quality standards empirically and can be used as an evaluation tool to measure concept understanding and detect student misconceptions in kinematics material. In terms of application, this instrument provides practical benefits for educators in developing learning strategies that are more suited to student needs. Theoretically, this research contributes to the development of diagnostic assessments based on Rasch Model analysis. However, this study has limitations in the limited number of respondents, namely only 30 students from one school, so it does not reflect the population as a whole. Therefore, it is recommended that further research be conducted with a wider and varied sample coverage to ensure the consistency of the results and strengthen the external validity of the instrument developed.

ACKNOWLEDGMENTS

We would like to express our deepest gratitude to Mrs. Yuvita Oktarisa, Ph.D, as the Head of Physics Education Department, who has provided guidance, direction, and motivation so that this report can be completed properly. Our gratitude also goes to Mr. Tresna Galih Sukma Suryana, M.Pd, as the Lecturer of Physics Learning Evaluation Course, who has provided guidance and direction so that this report can be completed on time. In, we would like to thank SMAN 4 Kota Serang for providing opportunities and facilities to conduct research. We would also like to thank our family and friends who always provide moral and material support.

REFERENCES

[1] A. H. Aldahmash and F. S. Alshaya, "Secondary School Student Alternative Conceptions about Genetics," Electronic Journal of Science Education, vol. 16, no. 1, pp. 1–21, 2012.
[2] N. F. Amin, S. Garancang, and K. Abunawas, "General Concepts of Population and Sample in the Study," PILAR Journal: Journal of Contemporary Islamic Studies, vol. 14, no. 1, 2023.
[3] A. Anastasi, Psychological Testing, 6th ed. New York, NY, USA: Macmillan, 1988.
[4] S. Arikunto, Assessment Procedures: A Practical Approach. Jakarta, Indonesia: Rineka Cipta, 1998.
[5] J. Day and D. Bonn, "Development of the concise data processing assessment," Physical Review Special Topics - Physics Education Research, vol. 7, no. 1, p. 10114, 2011. https://doi.org/10.1103/PhysRevSTPER.7.010114
[6] J. P. Guilford, Fundamental Statistics in Psychology and Education, 4th ed. New York, NY, USA: McGraw-Hill, 1956.
[7] O. P. U. Gumay, "Analysis of Student Misconceptions on Motion Material," Silampari Journal of Physical Science Education, vol. 3, no. 1, 2021. https://doi.org/10.31540/sjpif.v3i1.1239
[8] N. Hidayati, "The Use of Instruments in Educational Research," Journal of Education and Learning, vol. 23, no. 2, pp. 112–120, 2018.
[9] I. Kaniawati, N. J. Fratiwi, A. Danawan, I. Suyana, A. Samsudin, and E. Suhendi, "Misconceptions and Concept Change in Physics Education," Journal of Turkish Science Education, vol. 16, no. 1, 2019.
[10] M. Mustari, S. Anggereni, F. Sodikin, and A. D. Yusandika, "Identification of students' misconceptions using the Certainty of Response Index (CRI) from work and energy material," in Journal of Physics: Conference Series, vol. 1572, no. 1, 2020. https://doi.org/10.1088/1742-6596/1572/1/012038
[11] R. F. Nikat, A. Algiranto, M. Loupatty, and A. Henukh, "Conceptualizing Dynamics and Kinematics," Indonesian Journal of Science Education, vol. 10, no. 2, pp. 218–230, 2022.

https://doi.org/10.24815/jpsi.v10i2.23418
[12] N. Nurhadi, "The Effectiveness of Inquiry-Based Learning Methods on Kinematics Concept Understanding," Indonesian Journal of Physics Education, vol. 16, no. 2, pp. 123–130, 2020.
[13] I. Putra, E. Sujarwanto, and A. Sekar, "Analysis of Students' Conceptual Understanding of Particle Kinematics Material through Diagnostic Tests," Journal of Physics Education Research and Studies, vol. 5, no. 1, pp. 10–16, 2018.
[14] S. Azwar, Reliability, Validity, Interpretation and Computation. Yogyakarta, Indonesia: Liberty, 2006.
[15] R. Sari and A. Setiawan, "The Use of Computer Simulation Media in Kinematics Learning to Improve Student Understanding," Scientific Journal of Physics Education Al-Biruni, vol. 9, no. 1, pp. 45–52, 2021.
[16] S. Sugiono, Quantitative, Qualitative, and R&D Research Methods. Bandung, Indonesia: Alfabeta, 2016.
[17] M. Sugiyono, Quantitative, Qualitative, and Combination Research Methods. Bandung, Indonesia: Alfabeta, 2012.
[18] P. Suparno, Misconceptions and Concept Change in Physics Education. Jakarta, Indonesia: PT Grasindo, 2013.
[19] S. Suryabrata, Evaluation of Learning. Jakarta, Indonesia: PT RajaGrafindo Persada, 2010.
[20] M. Triastutik, A. Budiyono, and I. Diraya, "Identification of Student Misconceptions," Journal of Physics Innovation and Learning, vol. 8, no. 1, pp. 61–72, 2021. https://doi.org/10.36706/jipf.v8i1.13533
[21] C. Tuysuz, "Development of two-tier diagnostic instruments and a performance test for high school students' understanding of electrochemistry," Journal of Turkish Science Education, vol. 6, no. 2, pp. 88–101, 2009.
[22] Kutluay, Y. “Diagnosis of Eleventh Grade Students' Misconceptions about Geometric Optic by A Three Tier Test,” Turkey: Middle East Technical University. 2005