Journal of Innovative Physics Education Research (JIPER)
e-ISSN: 3089-5006 *Vol. 1 No. 1, April 2025, Page 1-14*

Development and Validation of Diagnostic Instrument to Identify Student Misconceptions in Vector Material

Arrafikar Rizqia Firdaus*, Elza Wulandarie, Mira Cantika, Tresna Galih Sukma Suryana
Department of Physics Education, Universitas Sultan Ageng Tirtayasa
Corresponding Author:** ***2280220044@untirta.ac.id

Received January 10, 2025; Revised January 30, 2025; Accepted March 26, 2025; Available online March 26, 2025
DOI: https://doi.org/10.61142/jiper.v1i1.195

Keywords *:*
Misconceptions, Vectors, Diagnostic Instrument, Physics Learning Evaluation

ABSTRACT
This study aims to develop and validate a web-based diagnostic instrument to detect students' misconceptions on vector material, including the concepts of vector magnitude, vector components, resultant, and vector displacement. The research was conducted using a quantitative research design, involving 30 students of class XI IPA 3 SMA Negeri 1 Padarincang. The research process includes instrument development, expert validation, limited trial, and data analysis. The results showed that the instrument had adequate reliability and validity, with KR-20 reliability of 0.848 (high category). The instrument is effective in identifying misconceptions, especially on the concepts of resultant and vector displacement. This article contributes to the development of diagnostic-based physics learning evaluation.

INTRODUCTION

Physics is a fundamental branch of natural sciences that explores and explains natural phenomena using systematic scientific processes [1]. These processes produce universally accepted principles, theories, and concepts that form the foundation for understanding the physical world. Among these foundational concepts, vectors are critically important as they are instrumental in understanding many advanced topics in physics, such as motion, forces, and energy. The ability to understand and apply vector concepts effectively is a prerequisite for mastering subsequent physics topics [2].

Despite its foundational importance, students often face difficulties in comprehending vector concepts [3]. Research has consistently shown that these difficulties stem from a combination of factors, including the abstract nature of vector representation, insufficient emphasis on conceptual understanding in traditional teaching methods, and limited opportunities for students to engage in hands-on learning activities [4]. These challenges lead to persistent misconceptions that hinder students' ability to build a solid foundation for further learning in physics.

Misconceptions in vector material can manifest in various ways, such as misunderstanding the difference between scalar and vector quantities, confusion in interpreting graphical representations of vectors, and errors in calculating resultant vectors [3]. These misconceptions are not merely errors but are deeply ingrained misunderstandings that conflict with scientifically accepted principles [5]. Addressing these misconceptions requires more than traditional assessment methods; it necessitates diagnostic tools that can accurately identify specific areas of misunderstanding and provide insights for targeted interventions [6].

The term "misconception" refers to a deviation in understanding from scientifically accurate concepts [7]. Such deviations are common in physics education and pose significant barriers to effective learning. If left unaddressed, misconceptions can perpetuate and become obstacles to mastering more complex topics [8]. For instance, students who fail to understand the concept of vector addition may struggle with understanding forces in equilibrium, limiting their ability to solve real-world problems involving multiple forces.

Given the critical need to address these issues, this study focuses on the development and validation of a diagnostic instrument specifically designed to identify misconceptions in vector material. Unlike traditional tests, which primarily evaluate students' ability to recall information, diagnostic instruments aim to probe deeper into students' reasoning processes [9]. By identifying the root causes of their misconceptions, educators can design instructional strategies that directly target these misunderstandings, thereby enhancing the overall effectiveness of teaching and learning [10].

The instrument developed in this study utilizes a web-based platform, leveraging the accessibility and flexibility of digital tools to reach a wider audience. The use of web-based diagnostics aligns with current trends in education, where technology is increasingly integrated into teaching and assessment practices. Web-based tools not only facilitate efficient data collection and analysis but also provide interactive and engaging experiences for students, making the diagnostic process more effective [11].

Robust statistical analysis methods are required to ensure the reliability and validity of diagnostic instruments [12]. The Rasch model, a probabilistic measurement approach based on Item Response Theory (IRT), is particularly useful in analyzing assessment data. The Rasch model provides a better understanding of how well each test item functions in differentiating student proficiency levels [13]. Unlike classical test theory, which assumes equal weighting for all items, the Rasch Model evaluates item difficulty, discrimination power and student ability on the same scale. This allows for a more precise measurement of conceptual understanding and identifies weak questions that may need to be revised or eliminated [14].

The specific objectives of this study are threefold: (1) to design a diagnostic instrument that accurately identifies misconceptions in vector material, (2) to validate the instrument through expert reviews and empirical trials, and (3) to analyze its reliability and effectiveness in diverse educational settings. By achieving these objectives, this research aims to contribute to the field of physics education by providing educators with a practical and scientifically grounded tool to improve student learning outcomes.

In summary, this study addresses a critical gap in physics education by focusing on the identification and remediation of misconceptions in vector material. The findings are expected to provide valuable insights for educators, researchers, and policymakers, highlighting the importance of diagnostic assessments in fostering a deeper understanding of fundamental physics concepts [15]. By leveraging the power of technology and innovative assessment methods, this research aspires to advance the quality of physics education and support students in developing a robust conceptual foundation.

METHOD

This study adopted a quantitative research approach with a developmental design. The research process was structured into three comprehensive stages to ensure the validity and reliability of the developed diagnostic instrument.

1. Preparation Stage

The development of diagnostic questions was based on the two-tier multiple-choice model. This model was chosen to allow for a deeper exploration of students' reasoning processes by requiring them to not only select an answer but also provide a justification. The initial draft of the questions was formulated based on common misconceptions identified in previous studies and expert recommendations. The questions were then reviewed for content relevance and clarity before proceeding to expert validation.

2. Implementation Stage

2.1 Expert Validation

To ensure content and construct validity, the instrument was reviewed by experienced physics education lecturers. They assessed the instrument based on predefined criteria, including question clarity, relevance to vector concepts, and effectiveness in diagnosing misconceptions. The feedback provided was used to refine the questions before testing them with student.

2.2 Limited Trials

A pilot study was conducted with 30 students from the XI Science 3 class at SMA Negeri 1 Padarincang. The trial was performed under normal classroom conditions to replicate real-world testing scenarios. The purpose of this phase was to determine whether the questions effectively identified misconceptions and whether students could understand and respond to them appropriately.

2.3 Data Collected

Responses from students were collected and analyzed using Microsoft Excel and Ministep software. Microsoft Excel was utilized for basic statistical calculations, including response frequency and item difficulty distribution, while Ministep was employed for Rasch model analysis to evaluate item reliability and validity.

3. Analysis and Refinement Stage

A comprehensive evaluation of the instrument was conducted to determine its validity, reliability, difficulty level, discrimination power, and distractor effectiveness .

3.1 Validity Analysis

Item validity was assessed using correlation analysis to determine how well each question measured the intended concept.

3.2 Reliability Analysis

The KR-20 formula was applied to measure internal consistency, ensuring that the instrument produced consistent results across different administrations.

3.3 Item Difficult and Discrimination Power

Difficulty levels were classified based on the proportion of correct responses, while discrimination indices were calculated to assess how well each question differentiated between high- and low-performing students.

3.4 Distractors Effectiveness

Each multiple-choice option was analysed to ensure that incorrect answers functioned as effective distractors by attracting a reasonable proportion of students. Based on the analysis, necessary refinements were implemented to improve the quality of the diagnostic instrument. Questions with low discrimination power or non-functional distractors were revised or replaced to enhance the instrument's overall effectiveness in identifying student misconceptions.

RESULTS AND DISCUSSIONS

1. Analysis of Difficulty Level

Item difficulty analysis aims to evaluate and improve the quality of items so that they can be used effectively in measuring the competence or knowledge of test takers accurately and fairly. Item difficulty levels are grouped into three main categories: easy, medium, and difficult. Questions are said to be easy if more than 70% of students answer correctly, moderate if between 30%-70% of students answer correctly, and difficult if less than 30% of students answer correctly. The following results of the analysis of difficulty level are presented using Ministep and manually Microsoft Excel.

INSTRUMENT TEST RECAPITULATION

No	Name	Question Number										Σ	C	W	N	UG-BG
		1	2	3	4	5	6	7	8	9	10					
	Σ (UG+MG+BG)	12	14	13	19	17	7	12	12	11	6					
	Difficulty Level (DL)	0.4	0.47	0.43	0.63	0.57	0.23	0.4	0.4	0.37	0.2					
	DL Categorization	MED	MED	MED	MED	MED	DIF	MED	MED	MED	DIF					
	Discriminating Power (DP)	0.2	0.2	0.9	0.6	0.4	0.2	0.2	0.5	0.5	0.5					
	DP Categorization	POS	POS	POS	POS	POS	POS	POS	POS	POS	POS					
Distarctors	A	0.17	0.1	0.2		0.2	0.43	0.17	0.23		0.13					
	B	0.2	0.17		0.13	0.07	0.17			0.13	0.13					
	C	0.13	0.03	0.13	0.1			0.2	0.2	0.33	0.27					
	D		0.23	0.1	0.1	0.1	0.1	0.1	0.03	0.07						
	E	0.1		0.13	0.03	0.07	0.07	0.13	0.13	0.1	0.27					

Description :
EA : Easy
DIF : Difficult
MED : Medium
CD : Can't Discriminte
POS : Positif
NEG : Negatif
UG : Upper Group
BG : Bottom Group
MG : Middle Group
Black Block : Answer Key
Red Block : Bad Distractors

***Fig 1**. Manual Level of Difficulty Data Using Microsoft Excel*

Figure 1 shows the difficulty level analyzed by calculating the proportion of correct answers for each question. Eight questions were categorized as moderate (index 0.3-0.7), providing an appropriate level of challenge for most students. Two questions were considered difficult (index < 0.3), indicating areas where students faced significant conceptual challenges, such as interpreting vector components and resultant calculations. These findings are critical for identifying gaps in student understanding.

Item STATISTICS: MEASURE ORDER

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
10	6	30	1.21	.50	.66	-1.20	.59	-.90	.68	.41	90.0	82.6	S10
6	7	30	.97	.48	1.35	1.25	1.46	1.16	.12	.42	73.3	80.4	S6
9	11	30	.18	.42	1.02	.15	.99	.04	.42	.43	70.0	71.9	S9
1	12	30	.01	.41	1.13	.79	1.15	.70	.31	.42	66.7	69.8	S1
7	12	30	.01	.41	1.02	.17	1.01	.11	.41	.42	73.3	69.8	S7
8	12	30	.01	.41	.95	-.23	.84	-.64	.49	.42	66.7	69.8	S8
3	13	30	-.16	.41	.70	-2.11	.66	-1.68	.68	.42	90.0	68.6	S3
2	14	30	-.32	.40	1.07	.54	1.01	.11	.36	.41	66.7	67.4	S2
5	17	30	-.80	.40	.97	-.15	1.09	.43	.39	.39	66.7	67.0	S5
4	19	30	-1.13	.41	1.00	.05	1.87	2.45	.28	.37	73.3	68.9	S4
MEAN	12.3	30.0	.00	.43	.99	-.07	1.07	.18			73.7	71.6	
P.SD	3.7	.0	.67	.03	.19	.92	.36	1.09			8.6	5.1	

***Fig 2**. Level of Difficulty Data Using Ministep*

In Figure 2, question number 10 is a difficult question, because out of 30 students only 6 people answered correctly and question number 6 out of 30 students only 7 people answered correctly, and question number 4 is an easy question because out of 30 students there are 19 people who answered correctly.

2. Analysis of Discriminating Power

Question discriminating power is the ability of a question to distinguish between high-ability students and low-ability students. The classification of distinguishing power is determined based on the discrimination index number (D) of the item. In other words, if an item has good discriminating power, it means that the item is able to distinguish between high-ability trainees and low-ability trainees. There are Differentiating Power Categories according to the magnitude of the D Value.

Table 1. Discriminating Power Category

Magnitude of the D Value	Discriminating Power Category
$D \leq 0$	Very Low
$0 < D \leq 0.2$	Low
$0.2 < D \leq 0.4$	Medium
$0.4 < D \leq 0.7$	High
$0.7 < D \leq 1$	Very High

Below are the results of the discriminating power analysis presented using Ministep and manually using Microsoft Excel.

INSTRUMENT TEST RECAPITULATION

No	Name	Question Number 1	2	3	4	5	6	7	8	9	10	∑	C	W	N	UG-BG
	∑ (UG+MG+BG)	12	14	13	19	17	7	12	12	11	6					
	Difficulty Level (DL)	0.4	0.47	0.43	0.63	0.57	0.23	0.4	0.4	0.37	0.2					
	DL Categorization	MED	MED	MED	MED	MED	DIF	MED	MED	MED	DIF					
	Discriminating Power (DP)	0.2	0.2	0.9	0.6	0.4	0.2	0.2	0.5	0.5	0.5					
	DP Categorization	POS	POS	POS	POS	POS	POS	POS	POS	POS	POS					
Distarctors	A	0.17	0.1	0.2		0.2	0.43	0.17	0.23		0.13					
	B	0.2	0.17		0.13	0.07	0.17			0.13	0.13					
	C	0.13	0.03	0.13	0.1			0.2	0.2	0.33	0.27					
	D		0.23	0.1	0.1	0.1	0.1	0.1	0.03	0.07						
	E	0.1		0.13	0.03	0.07	0.07	0.13	0.13	0.1	0.27					

Description :
EA : Easy
DIF : Difficult
MED : Medium
CD : Can't Discriminte
POS : Positif
NEG : Negatif
UG : Upper Group
BG : Bottom Group
MG : Middle Group
Black Block : Answer Key
Red Block : Bad Distractors

***Fig 3**. Manual Discriminating Power Data Using Microsoft Excel*

Based on Figure 3, questions number 1, 2, 6, and 7 have a discriminating power value of 0.2 which is included in the medium discriminating power category. Then in questions number 4, 5, 8, 9 and 10, including the high category because it has a value range of 0.4 <0.7, and finally in question number 3 has a value of 0.9, and is included in the Very high category. The discrimination index analysis revealed that 90% of the questions had positive values ranging from 0.2 to 0.9. This suggests that the questions effectively differentiated between high-performing and low-performing students. Questions with higher discrimination indices were particularly effective in pinpointing misconceptions, making them valuable for targeted instructional interventions.

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	19	30	-1.13	.41	1.00	.05	1.87	2.45	A .28	.37	73.3	68.9	S4
6	7	30	.97	.48	1.35	1.25	1.46	1.16	B .12	.42	73.3	80.4	S6
1	12	30	.01	.41	1.13	.79	1.15	.70	C .31	.42	66.7	69.8	S1
5	17	30	-.80	.40	.97	-.15	1.09	.43	D .39	.39	66.7	67.0	S5
2	14	30	-.32	.40	1.07	.54	1.01	.11	E .36	.41	66.7	67.4	S2
7	12	30	.01	.41	1.02	.17	1.01	.11	e .41	.42	73.3	69.8	S7
9	11	30	.18	.42	1.02	.15	.99	.04	d .42	.43	70.0	71.9	S9
8	12	30	.01	.41	.95	-.23	.84	-.64	c .49	.42	66.7	69.8	S8
3	13	30	-.16	.41	.70	-2.11	.66	-1.68	b .68	.42	90.0	68.6	S3
10	6	30	1.21	.50	.66	-1.20	.59	-.90	a .68	.41	90.0	82.6	S10
MEAN	12.3	30.0	.00	.43	.99	-.07	1.07	.18			73.7	71.6	
P.SD	3.7	.0	.67	.03	.19	.92	.36	1.09			8.6	5.1	

***Fig 4**. Discriminating Power Data Using Ministep*

The discriminating power analysis from the table above can be seen from the MNSQ values in INFIT and OUTFIT, both of which are positive, which means that the discriminating power of the question is accepted that the question can distinguish between high-ability students and low-ability students.

3. Analysis of Reliability

Reliability analysis is used to determine the consistency of the measuring instrument, whether the measuring instrument used is reliable and remains consistent if the measurement is repeated. Reliability shows how much consistency the test instrument has in measuring student abilities. Manual reliability analysis using Ms Excell can be used for dichotomous item analysis. On items with dichotomy, or having two answers, namely 1-0, true-false, yes-no, on-off, and others, it can be done using the KR-20 formula (Kuder Richardson 20). The KR-20 formula is as follows:

$$r_{11} = \frac{k}{k-1}\left(\frac{s^2 - \Sigma pq}{s^2}\right)$$

$$p = \frac{Np}{N};\ q = 1 - p \tag{1}$$

In general, reliability values are interpreted as follows:

Table 2. Reliability Interpretation Criteria

Magnitude of the Reliabilty Value	Reability Interpretation Category
0.9 - higher	Very High
0.80 – 0.89	High
0.70 – 0.79	Fair
0.60 – 0.69	Medium
Under 0.60	Low

The results of the reliability analysis of each question using the Ministep method and manually using Microsoft Excel have been carried out as follows.

RELIABILITY TEST

No	Name	Question Number										
		1	2	3	4	5	6	7	8	9	10	
1	Kevin Ndruru	0	0	0	0	0	1	1	0	0	0	
2	Dewi Puspita	1	0	0	0	0	0	0	0	1	0	
3	Ismawati	0	0	0	1	0	1	0	0	1	0	
4	Astuti Meilani	0	0	1	0	0	0	1	0	0	0	
5	Suci	1	1	0	0	0	0	0	0	0	0	
6	ELVA NAZILA	1	1	0	0	0	0	0	0	0	0	
7	Iyat mudiyati	1	0	0	1	0	0	1	1	0	0	
8	Arika	0	1	0	0	0	1	0	1	0	0	
9	nazwa solihah	0	0	0	1	0	0	0	0	0	0	
10	Annisya Yuniarti Qwarti	0	1	0	1	0	1	0	1	0	0	
11	Septiyani	0	0	0	1	1	0	1	0	1	0	
12	GINA NABILA	0	0	0	1	1	0	0	0	0	0	
13	SITI NADIN FADILAH	0	0	1	1	1	0	1	1	0	0	
14	Eqi Pirdiansyah	0	0	1	1	1	0	1	1	0	1	
15	Fathur Choer	1	1	1	0	1	1	1	1	1	1	
16	Aril Saputra	1	1	1	1	0	0	0	0	1	0	
17	Aris Firdaus	1	1	1	1	1	0	1	1	0	1	
18	Liah Masliah	1	1	1	1	0	0	1	1	1	1	
19	Wiwin	1	1	1	1	1	0	0	0	1	0	
20	EVAN RAMADHAN	1	0	1	0	1	1	0	0	1	0	
21	Fadilah Tunisa	0	0	1	1	1	0	0	1	0	0	
22	Muhamad basit rizky	0	1	1	1	1	0	1	0	1	1	
23	Zahra Eka	0	0	0	1	1	0	0	1	1	0	
24	Muhamad khudri	0	0	1	1	0	0	0	1	1	0	
25	Saepul Bahri	0	0	0	1	1	0	0	0	0	1	
26	Novan Subhan	1	0	0	1	1	0	0	1	0	0	
27	Rino Febrian	0	1	1	1	1	1	0	0	0	0	
28	Fitria Noviyanti	0	1	0	0	1	0	1	0	0	0	
29	Hana Nadifah	1	1	0	0	1	0	0	0	0	0	
30	Khairul Azis	0	1	0	0	1	0	1	0	0	0	
	Np	12	14	13	19	17	7	12	12	11	6	
	P	0.40	0.47	0.43	0.63	0.57	0.23	0.40	0.40	0.37	0.20	
	q	0.60	0.53	0.57	0.37	0.43	0.77	0.60	0.60	0.63	0.80	∑pq
	pq	0.240	0.249	0.246	0.232	0.246	0.179	0.240	0.240	0.232	0.160	2.263
	S	10										
	r11	1.0115	Very High									

__Fig 5__. Manual Reliability Data Using Microsoft Excel

From the results of data analysis in Figure 5, using the KR-20 formula the reliability of the instrument is 1.011453 with a very high category. This shows that the questions tested have a high level of rebility correlation which means that the questions are reliable. Reliability analysis was conducted using the KR-20 formula, resulting in a score of 0.848.

This high score indicates a consistent performance of the instrument across different administrations. The reliability test also highlighted that the majority of questions maintained internal consistency, reflecting the robustness of the instrument in accurately diagnosing misconceptions.

```
TABLE 3.1 SEKOLAH                                      ZOU718WS.TXT  Jun 02 2024 06:32
INPUT: 30 Person  10 Item  REPORTED: 30 Person  10 Item  2 CATS MINISTEP 5.7.2.0
--------------------------------------------------------------------------------

   SUMMARY OF 30 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN       4.1      10.0        -.43     .75      1.00    .00   1.07    .05 |
|  SEM        .4        .0         .19     .02       .03    .12    .10    .14 |
| P.SD       2.0        .0        1.03     .10       .17    .62    .53    .77 |
| S.SD       2.0        .0        1.05     .11       .17    .63    .54    .78 |
| MAX.       9.0      10.0        2.37    1.08      1.32   1.31   3.42   1.78 |
| MIN.       1.0      10.0       -2.36     .66       .68  -1.15    .43  -1.15 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .79 TRUE SD     .66  SEPARATION   .84  Person RELIABILITY  .42 |
|MODEL RMSE    .76 TRUE SD     .70  SEPARATION   .92  Person RELIABILITY  .46 |
| S.E. OF Person MEAN = .19                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .46  SEM = 1.44
STANDARDIZED (50 ITEM) RELIABILITY = .81

   SUMMARY OF 10 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE    S.E.      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      12.3      30.0         .00     .43       .99   -.07   1.07    .18 |
|  SEM       1.2        .0         .22     .01       .06    .31    .12    .36 |
| P.SD       3.7        .0         .67     .03       .19    .92    .36   1.09 |
| S.SD       3.9        .0         .71     .04       .20    .97    .38   1.14 |
| MAX.      19.0      30.0        1.21     .50      1.35   1.25   1.87   2.45 |
| MIN.       6.0      30.0       -1.13     .40       .66  -2.11    .59  -1.68 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .44 TRUE SD     .50  SEPARATION  1.14  Item   RELIABILITY  .57 |
|MODEL RMSE    .43 TRUE SD     .52  SEPARATION  1.21  Item   RELIABILITY  .59 |
| S.E. OF Item MEAN = .22                                                     |
-------------------------------------------------------------------------------
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000
```

***Fig 6**. Reliability Data Using Ministep*

Person reliability and item reliability are categorized into several categories, namely, weak (<0.67), fair (0.67- 0.80), good (0.80 - 0.90), excellent (0.91-0.94), and excellent (>0.94). Meanwhile, the Cronbach alpha value is categorized into poor (<0.50), poor (0.50-0.60), fair (0.60 - 0.70), good (0.70-0.80), excellent (>0.80). Ministep provided additional reliability statistics, including a person reliability of 0.72 and item reliability of 0.84, both reflecting acceptable consistency in measurement.

4. Analysis of Item Validity

Validity is the extent to which the accuracy and accuracy of a measuring instrument in performing its function. The validity test is a test that serves to see whether a measuring instrument is valid or invalid. The validity test categories using Microsoft Excel are as follows:

Table 3. Validity Interpretation Criteria

Magnitude of the Validity Value	**Reability Interpretation Category**
0.00 – 0.20	Very Low
0.20 – 0.40	Low
0.41 – 0.60	Fair
0.61 – 0.80	High
0.80 – 1.00	Very High

The results of the item validity analysis are presented using Ministep and manually Microsoft Excel, as follows.

VALIDITY TEST

No	Name	Question Number									
		1	2	3	4	5	6	7	8	9	10
	Hit 1	264	318	621	273	339	69	354	414	357	492
	Hit 2	216	224	221	209	221	161	216	216	209	144
	Hit 3	3501	3501	3501	3501	3501	3501	3501	3501	3501	3501
	Hit 4	869.61	886	880	855	880	751	870	870	855	710
	rxy	0.3036	0.36	0.71	0.32	0.39	0.09	0.41	0.48	0.42	0.69
Categorization		L	L	H	L	L	VL	L	F	F	H
	B	0									
	VL	1									
	L	5									
	F	2									
	H	2									
	VH	0									

***Fig 7**. Manual Validity Data Using Microsoft Excel*

From the results of data analysis in Figure 7, it was found that 1 question was categorized as very low with an rxy result of 0.09 on question item no. 6, 5 questions were categorized as low with rxy results of 0.3036, 0.36, 0.32, 0.39, and 0, 41 on items number 1, 2, 4, 5 and 7, 2 questions are categorized as fair with rxy results of 0.476 and 0.417 on items number 8 and 9, and 2 questions are categorized as high with rxy results of 0.706 and 0.693 on items number 3 and 10.

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
4	19	30	-1.13	.41	1.00	.05	1.87	2.45	A .28	.37	73.3	68.9	S4
6	7	30	.97	.48	1.35	1.29	1.46	1.16	B .12	.42	73.3	80.4	S6
1	12	30	.01	.41	1.13	.79	1.15	.70	C .31	.42	66.7	69.8	S1
5	17	30	-.80	.40	.97	-.15	1.09	.43	D .39	.39	66.7	67.0	S5
2	14	30	-.32	.40	1.07	.54	1.01	.11	E .36	.41	66.7	67.4	S2
7	12	30	.01	.41	1.02	.17	1.01	.11	e .41	.42	73.3	69.8	S7
9	11	30	.18	.42	1.02	.15	.99	.04	d .42	.43	70.0	71.9	S9
8	12	30	.01	.41	.95	-.23	.84	-.64	c .49	.42	66.7	69.8	S8
3	13	30	-.16	.41	.70	-2.11	.66	-1.68	b .68	.42	90.0	68.6	S3
10	6	30	1.21	.50	.66	-1.20	.59	-.90	a .68	.41	90.0	82.6	S10
MEAN	12.3	30.0	.00	.43	.99	-.07	1.07	.18			73.7	71.6	
P.SD	3.7	.0	.67	.03	.19	.92	.36	1.09			8.6	5.1	

***Fig 8**. Validity Data Using Ministep*

Figure 8 shows that the overall MNSQ value generated is 1.07 which means it is accepted, but there is 1 item that does not meet the criteria, namely number 4 which has an MNSQ value of 1.87. Likewise, the ZSTD value can also be said to be valid because overall it is still within the accepted value range of 0.18, it's just that there is 1 item that does not meet the criteria, namely number 4 has a ZSTD value of 2.45.

5. Analysis of Distractors

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. The results of the analysis of distractors are presented using Ministep and manually Microsoft Excel, as follows.

INSTRUMENT TEST RECAPITULATION

No	Name	Question Number 1	2	3	4	5	6	7	8	9	10	∑	C	W	N	UG-BG
	∑ (UG+MG+BG)	12	14	13	19	17	7	12	12	11	6					
	Difficulty Level (DL)	0.4	0.47	0.43	0.63	0.57	0.23	0.4	0.4	0.37	0.2					
	DL Categorization	MED	MED	MED	MED	MED	DIF	MED	MED	MED	DIF					
	Discriminating Power (DP)	0.2	0.2	0.9	0.6	0.4	0.2	0.2	0.5	0.5	0.5					
	DP Categorization	POS	POS	POS	POS	POS	POS	POS	POS	POS	POS					
Distarctors	A	0.17	0.1	0.2		0.2	0.43	0.17	0.23		0.13					
	B	0.2	0.17		0.13	0.07	0.17			0.13	0.13					
	C	0.13	0.03	0.13	0.1			0.2	0.2	0.33	0.27					
	D		0.23	0.1	0.1	0.1	0.1	0.1	0.03	0.07						
	E	0.1		0.13	0.03	0.07	0.07	0.13	0.13	0.1	0.27					

Description :
EA : Easy
DIF : Difficult
MED : Medium
CD : Can't Discriminte
POS : Positif
NEG : Negatif
UG : Upper Group
BG : Bottom Group
MG : Middle Group
Black Block : Answer Key
Red Block : Bad Distractors

***Fig 9**. Manual Distractors Data Using Microsoft Excel*

Manually, distractors are said to function properly if they are selected by at least 5% of the students who take the test. The following is described for each item.

DOI: https://doi.org/10.61142/jiper.v1i1.195

Table 4. Table of Distractor Analysis for Each Question Item

Number of Item	Analysis of Distractors	Recommendation
1	All distractors were functional, as all options were selected by more than 5% of participants	-
2	Distractor functionality is dominant, but option C was only selected by 3% of participants	Option C is replaced
3	All distractors were functional, as all options were selected by more than 5% of participants	-
4	Distractors are functioning, but option E is only chosen by 3% of participants	Option E is replaced
5	All distractors are functional, as all options were selected by more than 5% of participants	-
6	All distractors are working, as all options are selected by more than 5% of participants	-
7	All distractors are working, as all options were selected by more than 5% of participants	-
8	Distractor functioning has dominated, option D was only chosen by 3% of participants	Option D is replaced
9	All distractors are functional, as all options are selected by more than 5% of participants	-
10	All distractors are functional, as all options are selected by more than 5% of participants	-

TABLE 13.3 SEKOLAH ZOU718WS.TXT Jun 02 2024 06:32
INPUT: 30 Person 10 Item REPORTED: 30 Person 10 Item 2 CATS MINISTEP 5.7.2.0

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: MEASURE ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	Item
10	B	0	4	13	-1.10	.44	.26	.3	.4	-.25	S10
	A	0	4	13	-.97	1.06	.61	.8	.7	-.21	
	E	0	8	27	-.74	.62	.24	.6	.7	-.18	
	C	0	8	27	-.58	.39	.15	.5	.7	-.09	
	D	1	6	20	.97	1.04	.46	.6	.6	.68	
6	B	0	5	17	-.75	1.30	.65	1.9	1.5	-.14	S6
	E	0	2	7	-.69	.24	.24	.5	.6	-.07	
	A	0	13	43	-.42	.96	.28	1.4	1.3	.01	
	D	0	3	10	-.31	.44	.31	.8	1.0	.04	
	C	1	7	23	-.21	1.17	.48	1.7	1.5	.12	
9	E	0	3	10	-.93	.00	.00	.5	.6	-.16	S9
	B	0	4	13	-.87	.66	.38	.8	.7	-.17	
	D	0	2	7	-.76	.75	.75	1.0	.9	-.08	
	C	0	10	33	-.66	1.03	.34	1.3	1.3	-.16	
	A	1	11	37	.13	1.07	.34	1.0	1.0	.42	
1	C	0	4	13	-.96	.38	.22	.6	.6	-.20	S1
	B	0	6	20	-.71	.48	.21	.8	.8	-.14	
	A	0	5	17	-.68	1.15	.57	1.5	1.3	-.11	
	E	0	3	10	-.31	.57	.40	1.3	1.3	.04	
	D	1	12	40	-.04	1.24	.37	1.2	1.3	.31	
7	A	0	5	17	-1.35	.64	.32	.5	.5	-.40	S7
	E	0	4	13	-.72	.56	.32	.9	.8	-.11	
	C	0	6	20	-.56	.65	.29	1.1	1.0	-.06	
	D	0	3	10	-.30	.21	.15	1.1	1.1	.04	
	B	1	12	40	.08	1.22	.37	1.1	1.1	.41	
8	E	0	4	13	-1.35	.85	.49	.7	.6	-.35	S8
	D	0	1	3	-.93	.00		.6	.6	-.09	
	A	0	7	23	-.93	.85	.35	1.0	.9	-.27	
	C	0	6	20	-.39	.56	.25	1.2	1.2	.02	
2	D	0	7	23	-1.05	.90	.37	1.0	.8	-.33	S2
	A	0	3	10	-.61	.23	.16	.9	.9	-.06	
	B	0	5	17	-.57	.50	.25	1.1	1.0	-.06	
	C	0	1	3	-.45	.00		1.0	1.0	.00	
	E	1	14	47	-.03	1.18	.33	1.1	1.1	.36	
5	E	0	2	7	-1.94	.43	.43	.3	.3	-.39	S5
	D	0	3	10	-1.51	.00	.00	.4	.4	-.35	
	B	0	2	7	-.98	.53	.53	.9	.7	-.14	
	A	0	6	20	-.21	.83	.37	1.7	2.2	.11	
	C	1	17	57	-.08	.96	.24	.9	.9	.39	
4	E	0	1	3	-1.51	.00		.5	.4	-.19	S4
	B	0	4	13	-1.37	.25	.14	.6	.5	-.36	
	C	0	3	10	-.93	.00	.00	.8	.7	-.16	
	D	0	3	10	.29	1.60	1.13	2.6	7.1	.23	
	A	1	19	63	-.21*	.92	.22	1.0	1.1	.28	

* Average ability does not ascend with category score

Fig 10. *Distractors Data Using Ministep*

After analyzing each item, it has been found that in figure 10 there are many items whose distractors work well. This can be seen from the Percentage Count in the "Data" Table of each question that shows the distractor value is more than 5%, which indicates that the distractor of each question works well. But there are also items whose distractors do not work well, especially in question number 8 option D, question number 2 option C and question number 4 option E which have a percentage of 3%, indicating that these items do not work well and may have to be changed or replaced.

These results validate the instrument's effectiveness in identifying and addressing misconceptions while underscoring the necessity of iterative refinements to optimize its functionality. The combination of quantitative analysis and expert feedback provides a solid foundation for further applications and research in physics education.

CONCLUSION AND SUGGESTION

The web-based diagnostic instrument developed in this study proves to be a valuable tool for identifying student misconceptions in vector material. With demonstrated validity and reliability, the instrument offers a practical solution for educators aiming to enhance physics education through targeted interventions. Future research should focus on expanding the application of this instrument across varied educational contexts and refining its components based on feedback. Additionally, integrating this tool with other innovative teaching methods may further amplify its impact on learning outcomes.

ACKNOWLEDGMENTS

The authors express their sincere gratitude to SMA Negeri 1 Padarincang for their support and participation in this research. Special thanks are extended to Bapak Tresna Galih Sukma Suryana, M.Pd., and Ibu Yuvita Oktarisa, Ph.D., for their invaluable guidance and encouragement throughout the study. Their contributions have been instrumental in the successful completion of this research.

REFERENCES

[1] M. Panjaitan, H. Pardede, H. S. Handayani, and E. K. Sitinjak, "Analisis Miskonsepsi Buku Ajar Fisika SMA Kelas X Pada Materi Vektor," *J. Pendidik. dan Konseling*, vol. 2, no. 1, pp. 180–193, 2021. [Online]. Available: http://journal.universitaspahlawan.ac.id/index.php/jpdk/article/view/21301

[2] J. Wells, R. Henderson, A. Traxler, P. Miller, and J. Stewart, "Exploring the structure of misconceptions in the Force and Motion Conceptual Evaluation with modified module analysis," *Phys. Rev. Phys. Educ. Res.*, vol. 16, no. 1, p. 010121, Apr. 2020. Available: https://doi.org/10.1103/PhysRevPhysEducRes.16.010121

[3] S. Rachmasari, A. Prayitno, and S. Rahayuningsih, "Upaya Mengatasi Miskonsepsi Siswa pada Penjumlahan Vektor dengan Problem Based Learning," *Equal. J. Ilm. Pendidik. Mat.*, vol. 7, no. 1, pp. 47–54, 2024. Available: https://doi.org/10.46918/equals.v7i1.2131

[4] V. Nekhubvi and R. R. Ronella, "Analysis of the understanding of vectors in Physics for first-year university entering students," in *Proc. SAIP2023: 67th Annu. Conf. South African Inst. Phys.*, 2023, pp. 388–393.

[5] M. Lengkong, E. Istiyono, B. Rampean, A. Tumanggor, and M. Nirmala, "Development of Two-Tier Test Instruments to Detect Student's Physics Misconception," in *Proc. 7th Int. Conf. Res. Implementation Educ. Math. Sci. (ICRIEMS 2020)*, 2021. Available: https://doi.org/10.2991/assehr.k.210305.082

[6] D. Mutmainna, S. Mania, and A. Sriyanti, "Pengembangan Instrumen Tes Diagnostik Pilihan Ganda Dua Tingkat Untuk Mengidentifikasi Pemahaman Konsep Matematika," *MaPan*, vol. 6, no. 1, pp. 56–69, 2018, Available: https://doi.org/10.24252/mapan.2018v6n1a6

[7] J. Stewart et al., "Examining the relation of correct knowledge and misconceptions

using the nominal response model," *Phys. Rev. Phys. Educ. Res.*, vol. 17, no. 1, p. 010122, Mar. 2021. Available: https://doi.org/10.1103/PhysRevPhysEducRes.17.010122

[8] A. Myanda, M. P. Riezky, and M. Maridi, "Development of Two-Tier Multiple-Choice Test to Assess Students' Conceptual Understanding on Respiratory System Material of 11th Grade of Senior High School," *Int. J. Sci. Appl. Sci. Conf. Ser.*, vol. 4, no. 1, p. 44, 2020. Available: https://doi.org/10.20961/ijsascs.v4i1.49457.

[9] N. A. A. Sholihah, Sarwanto, and N. S. Aminah, "Development of two-tier multiple choice instrument to measure science process skill," *ISEJ Indones. Sci. Educ. J.*, vol. 1, no. 2, pp. 257–268, 2020. Available: https://doi.org/10.1088/1742-6596/1521/2/022053

[10] R. A. A. Ningroom, S. Yamtinah, and Riyadi, "A two-tier multiple-choice diagnostic test to find student misconceptions about the change of matter," *J. Educ. Learn.*, vol. 19, no. 2, pp. 1144–1156, 2025. Available: https://doi.org/10.11591/edulearn.v19i2.21478

[11] T. D. W. Ilahi, F. Mufit, H. Hidayati, and R. Afrizon, "Disain dan Validitas Multimedia Interaktif Berbasis Konflik Kognitif pada Materi Vektor untuk Kelas X SMA/MA," *J. Penelit. Pembelajaran Fis.*, vol. 12, no. 2, pp. 182–195, 2021. Available: https://doi.org/10.26877/jp2f.v12i2.9324

[12] M. Planinic, W. J. Boone, A. Susac, and L. Ivanjek, "Rasch analysis in physics education research: Why measurement matters," *Phys. Rev. Phys. Educ. Res.*, vol. 15, no. 2, p. 020111, Jul. 2019. Available: https://doi.org/10.1103/PhysRevPhysEducRes.15.020111

[13] D. R. Ocy, W. Rahayu, and M. Makmuri, "Rasch Model Analysis: Development of HOTS-Based Mathematical Abstraction Ability Instrument According to Riau Islands Culture," *AKSIOMA J. Progr. Stud. Pendidik. Mat.*, vol. 12, no. 4, p. 3542, 2023. Available: https://doi.org/10.24127/ajpm.v12i4.7613

[14] W. J. Boone, "Rasch Analysis for Instrument Development: Why, When, and How?," *CBE Life Sci. Educ.*, vol. 15, no. 4, 2016. Available: https://doi.org/10.1187/cbe.16-04-0148

[15] L. Ivanjek et al., "Development of a two-tier instrument on simple electric circuits," *Phys. Rev. Phys. Educ. Res.*, vol. 17, no. 2, p. 020123, Sep. 2021. Available: https://doi.org/10.1103/PhysRevPhysEducRes.17.020123