Development and Implementation of a Four-Tier Close-Ended Test to Analyze Students' Misconceptions of Optical Instruments

Itsna Rona Wahyu Astuti[1]*, Achmad Samsudin[1], Ida Kaniawati[1], Endi Suhendi[1], Bayram Çoştu[2]

[1]Department of Physics Education, Universitas Pendidikan Indonesia, Indonesia
[2]Program of Science Education, Yildiz Technical University, Turkey

*Correspondence Address:
itsnarona@upi.edu

Abstract: The research aims to develop a four-tier test for optical instrument materials. The method used in this study is a 4D design that includes defining, designing, developing, and disseminating. The instrument used consisted of fifteen items in the form of a four-tier closed-ended test. The research participants were 60 female and 15 male students from West Java in grade 11 high school who were randomly selected. The analysis is divided into four parts. The first analysis is a CVR and multi-rater Rasch measurement of the original validation results. The second analysis involves calculating the percentage of students' scores based on their conception scores. The third is a Rasch Model analysis of the instrument's validity and reliability. The Rasch Model is used in the fourth analysis to examine conceptions and misconceptions. Following the analysis, all items met the CVR value criteria. I2, I7, I9, I10, I12, I15, and I3 have logit values less than zero and are corrected based on expert feedback. The second analysis reveals that students continue to have misconceptions about each item. According to the third analysis, all items were valid and reliable, with a Cronbach Alpha value 0.78 in either category. According to the fourth analysis, conception is inversely related to misconception. The fewer misconceptions, the better the students' conceptions, and vice versa. However, confidence can also be a dissonant influence. Students who experience misconceptions need to be given appropriate treatment to reduce misconceptions about optical instrument materials. Hopefully, the four-tier closed-ended test that has been developed can be used and developed into a better five-level test to investigate the causes of each student's misconceptions.

INTRODUCTION

Misconceptions are the difference between student concepts and scientific theory concepts to understanding a phenomenon, and they are caused by competing beliefs that appear to be supported by logical arguments (Hammer, 1996; Heyd-Metzuyanim & Schwarz, 2017; Husnah et al., 2020; Jauhariyah et al., 2018; Kaltakci-Gurel et al., 2017; Kocakulah & Kural, 2010; Oberoi, 2017; Suprapto, 2020). Drawing on observations and practical experience, the concepts acquired over time tend to become ingrained in daily routines. Effecting a shift in students' misconceptions, moving them from incorrect to accurate, proves to be a challenging endeavor (Coştu et al., 2012; Greca et al., 2014; Kaltakci-Gurel et al., 2017; Liu & Fang, 2016; Ohlsson & Cosejo, 2014; Shen et al., 2017; Stein et al., 2008).

Misconceptions can be influenced by factors such as students' pre-existing knowledge, teachers' prior understanding, textbooks, the learning environment, and inaccuracies in transcribing terminology (Jauhariyah et al., 2018; Kocakulah & Kural, 2010). According to Oberoi (2017), Misconceptions can arise due to insufficient knowledge about concepts, textbook confusion, linguistic ambiguities, or overgeneralization. Misconceptions may also stem from learning strategies, students' initial information, the challenge of connecting one concept to another, the content presented in textbooks, and the influence of language and media. According to Oberoi (2017) and Kaltakçi & Didiç (2007), misconceptions among students can be caused by a variety of factors, including both the students themselves and their educational environment.

Misconceptions can be identified through the use of diagnostic tests, which are assessment tools designed to pinpoint challenges or unresolved issues that learners may encounter in the learning process (Fariyani et al., 2017; Gurel et al., 2015; Pertiwi & Setyarsih, 2015; Rosita et al., 2020). Diagnostic tests are also characterized as assessments designed to identify students' weaknesses, facilitating the implementation of appropriate measures to address those areas of concern (Ismail et al., 2015; Rosita et al., 2020). Interviews, concept maps, open tests, multiple-choice tests, two-tier multiple-choice tests, three-tier tests, and four-tier tests can all uncover students' misconceptions. Interviews, open-ended tests, and multiple-choice tests are commonly employed in physics education research. Nevertheless, each diagnostic test instrument comes with its own set of advantages and disadvantages when compared to others (Gurel et al., 2015; Kaltakci-Gurel et al., 2017).

Open-ended tests or descriptive assessments can prompt students to contemplate a concept for an extended period, articulate their thoughts in writing, uncover misconceptions in problem-solving, and assist students in overcoming learning difficulties. The open-ended test format enables respondents to express their answers in their own words and allows the administration of a broader sample than interview tests (Gurel et al., 2015; Kaltakci-Gurel et al., 2017; Zhou et al., 2016). Nevertheless, tests in the form of descriptive responses come with practical limitations related to language use issues. Students often exhibit a lack of enthusiasm in providing answers as complete sentences, necessitating more time for result analysis and assessment (Bautista & Boone, 2015; Kaltakci-Gurel et al., 2017; Kaltakçi & Didiç, 2007).

The four-tier test is structured with four levels. The first level involves answer choices (multiple-choice), the second level involves indicating the level of confidence for the answers selected in the first level, the third level entails choosing reasons (multiple-choice) for the answers selected in the first level, and the fourth level involves indicating the confidence level for the reasons selected in the third level (Kaltakci-Gurel et al., 2017). In the second and fourth levels, the expressions of confidence typically involve categorizations such as "sure" and "not sure." The four-tier test represents an advancement over a comparable diagnostic test with a three-tier format comprising only three components. The three-level diagnostic test, in turn, enhances a two-level diagnostic test. Incorporating reasons for selecting answers is a notable improvement (Anggrayni & Ermawati, 2019; Hermita et al., 2017). Therefore, the four-level multiple-choice diagnostic test is the most accurate for detecting misconceptions (Afif et al., 2017; Anggrayni & Ermawati, 2019; Hermita et al., 2017). The comprehension levels of the four-tier test reveal that students' conceptions can be categorized into six distinct conceptual levels according to the level of

understanding conveyed by Coştu (2008), the assessment by Kaltakci-Gurel et al. (2017), and the concept category of (Amalia et al., 2019). The six categories are Sound Understanding (SU), Partial Positive (PP), Partial Negative (PN), Misconception (MC), No Understanding (NU), and No Coding (NC).

Based on literature studies, students still have misconceptions about some physics concepts (Coetzee & Imenda, 2012; Kocakulah & Kural, 2010; Kucukozer, 2010; Rohmanasari & Ermawati, 2020; Salamah et al., 2017; Salmadhia et al., 2021; Umar et al., 2021). One of the physics materials that still has misconceptions is optical instrument material (Kaniawati et al., 2020; Rohmanasari & Ermawati, 2020; Salmadhia et al., 2021). Consequently, researchers necessitate a suitable instrument to evaluate students' misconceptions regarding optical

instruments. Accordingly, a four-tier closed-ended test focusing on optical devices has been devised, which deviates from the open-ended format in the third tier.

METHOD
Research Design

The research method followed the 4D model, encompassing the stages of defining, designing, developing, and disseminating (Thiagarajan et al., 1974). The defined stage of literature studies on optical instrument misconceptions has been concluded. The subsequent design stage involved establishing a construction distribution for each item, designing content for each item, and implementing a four-tier test. The first, second, and fourth tiers utilize closed-ended questions in this test, while the third tier adopts an open-ended format.
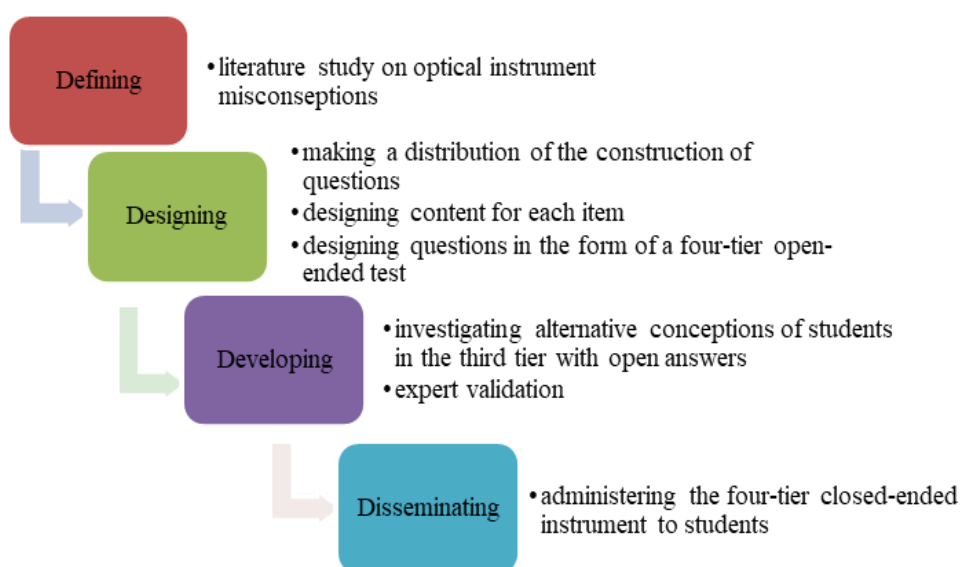
Figure 1. Research Chart of the Four-Tier Closed-Ended Questions on Optical Instruments.

In the development stage, the four-tier open-ended format is expanded to involve students explaining a concept in the third tier. Subsequently, all the reasons provided by the students in the third tier are compiled and used as answer choices. This restructuring transforms the four-tier open-

ended test into a four-tier closed-ended test. The instrument underwent validation by five experts who assessed each item based on nine validation indicators. The validation results were then analyzed using the Content Validity Ratio (CVR) and multifaceted Rasch measurement. In the

dissemination stage, the finalized four-tier closed-ended instrument, analyzed using the Rasch Model, was tested.

Participants

The participants involved in this study were 75 high school students in grade 11 in West Java. The students consisted of 60 female students and 15 male students. Random sampling is used to select students.

Instruments

The research employed a four-tier closed-ended instrument comprising fifteen items related to optical instruments, including topics such as cameras, eyes and eyeglasses, magnifying glasses, microscopes, and binoculars. The first tier presents a concept-based question, while the second tier gauges students' confidence in responding to the initial question. The third tier requires students to provide reasons or explanations for their answers to the first tier. Lastly, the fourth tier assesses the confidence level associated with the explanations provided in the third tier. All questions in the test are closed-ended and take the form of multiple-choice queries.

Data Analysis

The analysis conducted in this study comprises four stages. The expert validation results were initially scrutinized, employing CVR and multi-rater Rasch measurement for analysis. The experts evaluated the validity of the subject under consideration, determining whether it is valid without revision, with revision, or invalid. Equation 1 was applied to calculate CVR.

$$CVR = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}$$

Description:
n_e = The number of validators who provide valid assessments
N = The total number of validators

The instrument is deemed valid if the calculated CVR result surpasses the minimum CVR value, as determined by the Schipper Table (Wilson et al., 2012). Table 1 shows the minimum CVR values for the various validator counts.

Table 1. The Minimum CVR Values for the Various Validator Numbers.

Number of Experts	Minimum CVR Value	Number of Experts	Minimum CVR Value
5	0.736	15	0.425
6	0.672	20	0.368
7	0.622	25	0.329
8	0.582	30	0.300
9	0.548	35	0.287
10	0.520	40	0.260

The multi-rater Rasch measurement model is used to analyze the expert validation. The multi-rater Rasch measurement model created by Linacre is as follows (Eckes, 2019):

$$\ln\left[\frac{p_{nijk}}{p_{nijk-1}}\right] = \theta_n - \beta_i - \alpha_j - \tau_k$$

Description:
p_{nijk} = Probability of examinee n receiving a rating of k on criterion i from rater j

p_{nijk-1} = Probability of examinee n receiving a rating of k-1 on criterion i from rater j
θ_n = Proficiency of examinee
β_i = Difficulty of criterion i,
α_j = Severity of rater j,
τ_k = Difficulty receiving a k rating relative to a k – 1

The second step is categorizing students in each concept understanding category based on their responses. The

results of the classification of concept categories for students are expressed in percent form. The rubric of the conception category, conception score, and misconception score are shown in Table 2 (Amalia et al., 2019; Aminudin et al., 2019; Coştu, 2008; Kaltakci-Gurel et al., 2017).

Table 2. The Score of Conceptions and Misconception.

Categories	Tier				Score	MC
	1	**2**	**3**	**4**		
Sound Understanding (SU)	True	Sure	True	Sure	4	0
Partial Positive (PP)	True	Sure	True	Unsure		
	True	Unsure	True	Sure	3	0
	True	Unsure	True	Unsure		
Partial Negative (PN)	True	Sure	False	Sure		
	True	Sure	False	Unsure		
	True	Unsure	False	Sure		
	True	Unsure	False	Unsure	1	1
	False	Sure	True	Sure		
	False	Sure	True	Unsure		
	False	Unsure	True	Sure		
	False	Unsure	True	Unsure		
No Understanding (NU)	False	Unsure	False	Sure		
	False	Sure	False	Unsure	0	3
	False	Unsure	False	Unsure		
Misconception (MC)	False	Sure	False	Sure	0	4
No Coding (NC)	(Incomplete answer)				-	-

The third analysis stage evaluates the four-tier optical instrument questions developed using the Rasch Model:

$$P_{ni}\big(x_{ni} = 1|\beta_n, \delta_i\big) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}}$$

$P_{ni}\big(x_{ni} = 1|\beta_n, \delta_i\big)$ is the probability of respondent n in i to produce the correct answer $(x_{ni} = 1)$ with the respondent's ability (β_n) and item difficulty level (δ_i) (Sumintono & Widhiarsho, 2015).

The instruments were subjected to data analysis, focusing on students' conception scores for each item. Instrument analysis was employed to assess the items' validity, reliability, and difficulty level. Rasch analysis served as the methodology for instrument analysis. The instrument's validity is gauged by evaluating the appropriateness of each item. Item validity is determined through the output of the Item Fit Order, considering outfit mean square (MNSQ), outfit Z-Standard (ZSTD), and point measure correlation (PT MEASURE CORR). Additionally, the unidimensionality output, indicating the raw variance explained by measures, is utilized to ascertain instrument validity.

Rasch analysis was also employed to assess the reliability of the instruments, yielding results such as person reliability, item reliability, and Cronbach's Alpha. Person reliability gauges the consistency of students' responses, while item reliability reflects the quality of the instrument's items. Cronbach's Alpha provides an overview of the overall interaction between individuals and items.

In the fourth analysis, students' misconception scores on each item were scrutinized using Rasch analysis. Output tables, such as output variable maps (Wright maps), were utilized in this analysis to interpret the findings.

RESULT AND DISCUSSION

The focus is exploring alternative conceptions derived from students' responses in transitioning from the four-tier open-ended to the four-tier closed-ended tests on optical instruments. The subsequent sections will provide a detailed

discussion of the stages of development (define, design, develop, and disseminate) and the associated analysis within the framework of the 4D model.

Define

The define stage is a review of the literature on optical instrument misconceptions. This stage is utilized to locate research sources. A literature review of misconceptions about optical instruments is performed on each sub-material. Eyes, cameras, eyeglasses, magnifying glasses, microscopes, and binoculars are the optical instrument materials investigated. Based on the literature review and the prediction of optical instrument misconceptions that students may encounter a four-tier test can be developed. The results of literature studies, the misconceptions that occur among students regarding optical instruments are detailed as follows (Munawaroh et al., 2016; Kaniawati et al., 2020; Salmadhia, Rusnayati, & Liliawati, 2021).

Table 3. Students Misconceptions about Each Sub-material.

Sub Material	Students Misconceptions
Eyes, eyes glasses, and camera	The near point (PP) of hyperopia is farther than the normal eye, so objects must be placed closer than 25 cm. The larger the field diameter of a lens, the more light comes in, so the image gets bigger. The older you get, the better your eyes' accommodation power becomes. The pupil in the human eye has the same function as the diaphragm, regulating the intensity of incoming light. The near point of the eye of a myopia sufferer is closer than the near point of a normal eye (PP < 25cm). The camera's distance to the object is closer when the camera is in a landscape position than in a portrait position.
Magnifying glass	If the loupe lens is partially closed, the image formed is half of the object. A convex lens is used as a loupe because it spreads light so that the image of an object is enlarged from its original size. The strength of the loop is not affected by the medium in which the loop is used, meaning that the strength of the loop in air and water is the same.
Microscope	The function of a microscope is to see small objects so that they appear large and clear. Observations using a microscope with maximum accommodation occur when the image formed by the objective lens is exactly in the focus of the eyepiece lens. Concave mirrors and convex lenses have the property of scattering light.
Binoculars	All lenses on stage binoculars are convex lenses that can collect light. In unaccommodated observations, the magnification of the image from a star telescope is influenced by the length of the telescope tube. The more lenses a binocular has the greater the angular magnification.

Design

The instrument design stage requires distributing the construction of questions for each sub-material of optical instruments, designing content for each item, and designing questions in the form of a four-tier open-ended test. The first tier is a regular multiple choice, while the second tier is about confidence, with two options: "sure" and "unsure." However, students have no choice but to fill in at the third tier of the four-tier open-ended as in Figure 2a. At the same time, the fourth tier of confidence is comparable to the second tier. After obtaining alternative concepts from students' answers at the third tier, design a four-tier close-ended test as in Figure 2b.

Question 1.1 (Multiple choice questions with five answer choices and according to item construction). **Question 1.2** Are you sure about your answer to question 1.1? A. Sure B. Unsure **Question 1.3** Reasons for the answer to question 1.1: .... **Question 1.4** Are you sure about your answer to question 1.3? A. Sure B. Unsure	**Question 1.1** (Multiple choice questions with five answer choices and according to item construction). **Question 1.2** Are you sure about your answer to question 1.1? A. Sure B. Unsure **Question 1.3** Reasons for the answer to question 1.1: (There are five answer choice) **Question 1.4** Are you sure about your answer to question 1.3? A. Sure B. Unsure
(a)	**(b)**

Figure 2. The Design of Four-tier: (a) Open-ended Test, and (b) Close-ended Test

The distribution of the question construction contains the form of questions and choices in the first tier. Each set of questions and choices in the first tier can be in statements, pictures, or tables, such as the item construction distribution in Table 3. In the construction distribution of eye, camera, and glasses sub-materials, they are included in one sub-material because they work similarly. This sub-material also has the most questions compared to the sub-materials of magnifying glasses, microscopes, and binoculars. This aligns with (Kaniawati et al., 2020; Munawaroh et al., 2016) that the most common misconceptions are in the eye, camera, and glasses sub-materials.

Table 4. Distribution of Item Construction.

Sub Material	Item Number	Item Construction
Eyes, eyes glasses, and camera	1	Statement, Figure – Table
	2	Statement – Statement
	3	Table – Statement
	4	Figure – Table
	5	Statement – Table
	6	Statement, Figure – Statement
Magnifying glass	7	Statement, Figure – Figure
	8	Figure, Statement – Statement
	9	Table – Statement
Microscope	10	Figure – Statement
	11	Statement – Figure, Table
	12	Statement – Figure
Binoculars	13	Figure, Statement – Figure
	14	Table – Statement
	15	Statement, Figure – Statement

Develop

During the development stage, open-ended four-tier instruments will be converted to four-tier close-ended instruments. The four-tier open-ended instruments consist of 15 items. This stage was carried out to investigate alternative conceptions of students in the third tier with open answers. Furthermore, the alternative concept obtained will be an option on the third tier of the four-tier closed-ended instrument. Figure 3 shows an example of a third-tier change from open to closed.

Question 6.1
Ana wants to photograph an object in the form of flowers using a camera. When taking pictures, Ana tries two camera positions, namely landscape and portrait, with the same camera. Next, Ana gets a portrait image like Figure 3.

Figure 3(a). Photo results in landscape position

Figure 3 (b). Photo results in portrait position

Based on the photos taken by Ana, it shows the distance of the camera to the object when the camera is in a ....
A. landscape is closer than portrait
B. portrait is closer than landscape
C. portrait and landscape are the same
D. portrait and landscape are undefined
E. portrait and landscape are parallel

Question 6.2
Are you sure about your answer to question 6.1?
A. Sure
B. Unsure

Question 6.3
Reasons for answering question 6.1:
....
....
....

Question 6.4
Are you sure about your answer to question 6.4?
A. Sure
B. Unsure
Answer keys: 6.1: C

Question 6.1
Ana wants to photograph an object in the form of flowers using a camera. When taking pictures, Ana tries two camera positions, namely landscape and portrait, with the same camera. Next, Ana gets a portrait image like Figure 3.

Figure 3(a). Photo results in landscape position

Figure 3 (b). Photo results in portrait position

Based on the photos taken by Ana, it shows the distance of the camera to the object when the camera is in a ....
A. landscape is closer than portrait
B. portrait is closer than landscape
C. portrait and landscape are the same
D. portrait and landscape are undefined
E. portrait and landscape are parallel

Question 6.2
Are you sure about your answer to question 6.1?
A. Sure
B. Unsure

Question 6.3
Reasons for answering question 6.1:
A. because the landscape side has a shorter vertical side than the horizontal side
B. because on a portrait camera, the object it captures looks closer and doesn't contain too much background, so the camera's focus is more on the object
C. because portrait images produce images that look closer
D. because portrait and landscape cannot be determined
E. because the difference is that portrait is oriented vertically while the landscape is oriented horizontally. So, the distance of the camera to the object is the same

Question 6.4
Are you sure about your answer to question 6.4?
A. Sure
B. Unsure
Answer keys: 6.1: C and 6.3: E

Figure 3. The Example of the Four-tier Close-ended Test on Optical Instruments.

In the four-tier closed-ended test's Question 6.3, the answer choices represent reasons given by students. Each reason is categorized, modified, and used as a response option. This approach is highly beneficial for probing student responses. According to (Caleon & Subramaniam, 2010; Gurel et al., 2015; Kaltakci-Gurel et al., 2017), reasoning can differentiate between a correct answer due to the right rationale (scientific concept) and a correct answer based on flawed reasoning (false positive).

Table 5. Analysis Utilizing CVR.

Item	AAN	N	Ne	CVRi	CVRa
1	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
2	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
3	1	5	5	1	0.867
	2	5	4	0.6	
	3	5	4	0.6	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	4	0.6	
	9	5	5	1	
4	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
5	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
6	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
7	1	5	5	1	0.956
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	4	0.6	
	9	5	5	1	
8	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
9	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
10	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
11	1	5	5	1	0.956
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	4	0.6	
	9	5	5	1	
12	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
13	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
14	1	5	5	1	1.000
	2	5	5	1	
	3	5	5	1	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	
15	1	5	5	1	0.956
	2	5	5	1	
	3	5	4	0.6	
	4	5	5	1	
	5	5	5	1	
	6	5	5	1	
	7	5	5	1	
	8	5	5	1	
	9	5	5	1	

*AAN: Assessment Aspect Number; N: the total number of validators; Ne: the number of validators who provide valid; CVRi: CVR index; CVRa: CVR average.

```
+-------------------------------------------------------------------------------------------------------------+
|Measr|+Item    |-Assessment Aspect                                                              |-Expert|Scale|
|-----+---------+---------------------------------------------------------------------------------+-------+-----|
|  2 + I1       +                                                                                 +       + (3) |
|     |         |                                                                                 |       |     |
|     | I4  I5  I6 | the answer choices and reasons are homogeneous and logical in terms of material |    |     |
|   : | I8      :                                                                                 :       :     |
|  1 +         + questions do not provide clues to the correct answer                             +       + --- |
|     |         | items are made to determine the understanding of students' concepts            |       |     |
|   : |         : the consistency of the concepts in the questions with the concepts advanced by the experts : :  :
|     | I13     |                                                                                 |       |     |
|     |         | items are made according to misconceptions                                      |       |     |
|*  0 * I14     * there is only one answer key                                                    *       * 2 * |
|     | I2  I7  I9 |                                                                               |       |     |
|     | I10     | the language used is easy for students to understand                            |       |     |
|     | I12 I15 |                                                                                 |       |     |
| -1 + I11     +                                                                                 +       + --- |
|     |         |                                                                                 |       |     |
|     | I3      |                                                                                 |       |     |
|     |         | answer choices do not use statements all answers are correct or answers are wrong |     |     |
|   : |         : utilizing language that follows the rules of the Indonesian language            :       :     |
| -2 +         +                                                                                 +       +     |
|     |         |                                                                                 |       |     |
|     |         |                                                                                 |   D   |     |
| -3 +         +                                                                                 +       +     |
|     |         |                                                                                 |       |     |
|     |         |                                                                                 |   C   |     |
| -4 +         +                                                                                 +       +     |
|     |         |                                                                                 |   E   |     |
| -5 +         +                                                                                 + A  B  + (1) |
|-----+---------+---------------------------------------------------------------------------------+-------+-----|
|Measr|+Item    |-Assessment Aspect                                                              |-Expert|Scale|
+-------------------------------------------------------------------------------------------------------------+
```

Figure 4. Multi-rater Validation Test Results.

```
+----------------------------------------------------------------------------------------------------------+
| Total  Total  Obsvd  Fair(M)|   -    Model| Infit      Outfit     |Estim.| Correlation | Exact Agree. |        |
| Score  Count Average Average|Measure S.E. | MnSq ZStd  MnSq ZStd  |Discrm| PtMea PtExp | Obs % Exp % | N Expert|
|----------------------------------------------------------------------------------------------------------|
|  367    135   2.72   2.81  | -2.24  .19 | .72 -1.9   .50 -1.8  | 1.27 | .58   .46  | 73.7  75.3 | 4 D    |
|  394    135   2.92   2.95  | -3.74  .32 | 1.17  .6   4.13  2.9 | .86  | .16   .27  | 86.5  87.2 | 3 C    |
|  400    135   2.96   2.98  | -4.59  .46 | 1.06  .2   8.40  3.3 | .84  | .01   .18  | 89.1  89.5 | 5 E    |
|  405    135   3.00   3.00  |( -7.44 1.83)|Minimum            |      | .00   .00  | 91.1  91.1 | 1 A    |
|  405    135   3.00   3.00  |( -7.44 1.83)|Minimum            |      | .00   .00  | 91.1  91.1 | 2 B    |
|----------------------------------------------------------------------------------------------------------|
|  394.2  135.0  2.92   2.95  | -5.09  .93 | .98  -.3   4.34  1.5 |      | .15   |            | Mean (Count: 5)  |
|   14.2    .0    .11    .07  |  2.06  .75 | .19  1.1   3.23  2.4 |      | .22   |            | S.D. (Population) |
|   15.9    .0    .12    .08  |  2.30  .83 | .23  1.4   3.95  2.9 |      | .25   |            | S.D. (Sample)    |
+----------------------------------------------------------------------------------------------------------+

With extremes, Model, Populn: RMSE 1.19 Adj (True) S.D. 1.68 Separation 1.41 Strata 2.22 Reliability (not inter-rater) .67
With extremes, Model, Sample: RMSE 1.19 Adj (True) S.D. 1.97 Separation 1.66 Strata 2.54 Reliability (not inter-rater) .73
Without extremes, Model, Populn: RMSE .34 Adj (True) S.D. .91 Separation 2.67 Strata 3.89 Reliability (not inter-rater) .88
Without extremes, Model, Sample: RMSE .34 Adj (True) S.D. 1.14 Separation 3.34 Strata 4.79 Reliability (not inter-rater) .92
With extremes, Model, Fixed (all same) chi-squared: 44.8 d.f.: 4 significance (probability): .00
With extremes, Model, Random (normal) chi-squared: 4.3 d.f.: 3 significance (probability): .23
Inter-Rater agreement opportunities: 1350 Exact agreements: 1165 = 86.3% Expected: 1172.3 = 86.8%
```

Figure 5. Expert Measurement Report.

After converting all items to a four-tier closed-ended format, expert validation is conducted with input from five validators. The validators assessing the four-tier closed-ended test include three physics education lecturers, a teacher, and a researcher in the same field. The evaluation conducted by the validators encompasses nine aspects: 1) Items are crafted based on misconceptions; 2) Consistency of the concepts in the questions with those advanced by the experts; 3) Items are designed to assess the understanding of students' concepts; 4) Utilization of language that adheres to the rules of the Indonesian language; 5) Language used is accessible and understandable for students; 6) Answer choices and reasons exhibit homogeneity and logical alignment with the material; 7) There is only one correct answer key; 8) Questions do not provide hints or clues to the correct answer; 9) Answer choices do not include statements like "all answers are correct" or "answers are wrong." The results of the validation analysis utilizing CVR are presented in Table 5.

As per Table 5, all items exhibit an average CVR value greater than or equal

to 0.736. Given that the smallest CVR value with five validators is 0.736, it can be concluded that the questions are considered valid and can be utilized (Wilson et al., 2012). The average CVR calculation results for each item can be interpreted to mean that all items have valid expert validation results. However, when each item on each aspect is reviewed, the assessment reveals that the CVR values on items 3, 10, 11, and 15 need to be improved. It needs to be improved, according to assessment aspects 2, 3, and 8 in item 3. Meanwhile, items 10 and 11 only require refinement regarding judging questions that do not provide hints of the correct answer. Item 15 was revised in response to the comments on assessment aspect number 3.

The multi-rater Rasch measurement was utilized to analyze the results of expert validation. Figure 1 illustrates the outcomes of the multi-rater analysis, featuring five columns. The first column, known as the size column (logit transformation), displays measurement results with values ranging from +2 (top) to -5 (bottom), representing logit values. The second column delineates the distribution of logit values, spanning from less than logit -1 (I11) to greater than logit +2 (I1). The logit value of 0 serves as the minimum criterion for item quality, with experts considering values above this threshold as indicative of good-quality items and values below as representing items of lesser quality.

Figure 4 shows eight items considered unfavorable by experts: item numbers I2, I7, I9, I10, I12, I15, I11, and I3. Meanwhile, the expert deemed seven items qualified, including I14, I13, I8, I4, I5, I6, and I1. The third column in Figure 4 illustrates information about the difficulty level of the assessment aspects. This column displays the distribution of the assessment aspects. According to the experts, a lower logit value for an assessment aspect indicates that fulfilling an item in that aspect is easier. Conversely, a higher logit value suggests greater difficulty for the assessment aspect to be fulfilled in an item, as evaluated by the validator. Assessment aspects with similar logit values share the same level of difficulty.

According to Figure 4, assessment aspects 4 (using language that follows the rules of the Indonesian language) and 9 (answer choices do not use statements; all answers are correct or answers are wrong) are the easiest aspects of judging because all items satisfy this assessment aspect according to the validators. While assessment aspect 6 (the answer choices and reasons are homogeneous and logical in terms of material) is the most difficult aspect of the assessment, most question items do not meet it in the expert's opinion. According to all expert opinions, I1 is the only item that satisfies all aspects of the assessment.

Figure 5 depicts the quality of an expert panel's assessment, sorted by item severity. Expert D is the most consistent when considering statistical fit criteria (Boone et al., 2014). Expert D has Outfit MNSQ and Outfit ZSTD values in the statistical suitability ranges of 0.5-1.5 (Outfit MNSQ) and -2 to +2 (Outfit ZSTD), respectively (Outfit ZSTD). Experts A and B are the worst because they have the lowest infit value. The reliability between raters is sufficient (0.67), indicating that the experts give quite different scores, but some are the same (Koçak, 2020). The rater's tendency influences the reliability value (Bond & Fox, 2013). The obtained data aligns with the measurement model, and this alignment is corroborated by the Chi-square test value ($p < 0.01$). The agreement in assessment by the five experts (inter-rater agreement) stands at 86.3%, signifying minimal divergence in evaluating all items among the five experts. Existing studies consistently

indicate variations in raters' judgment tendencies, with rater behaviors such as leniency and severity influencing rater reliability (Brookhart et al., 2006; Darmana et al., 2021; Güler, 2014).

Disseminate

The disseminated stage is a concrete stage for putting the instrument for utilization. The results are then examined in three stages of analysis. The first analysis determines the percentage of each conception category derived from student scores. Thus, students' conceptions can be distributed based on the categories created. The second analysis analyzes the closed-ended four-tier instrument based on score conceptions. The third analysis includes a detailed description of conception and misconception and a comparison using Rasch analysis.

Based on the results, the conception score of all students on each item can be determined. The conception score per maximum conception value shows the percentage value of each conception. Figure 6 depicts the proportion of conception categories for each item.
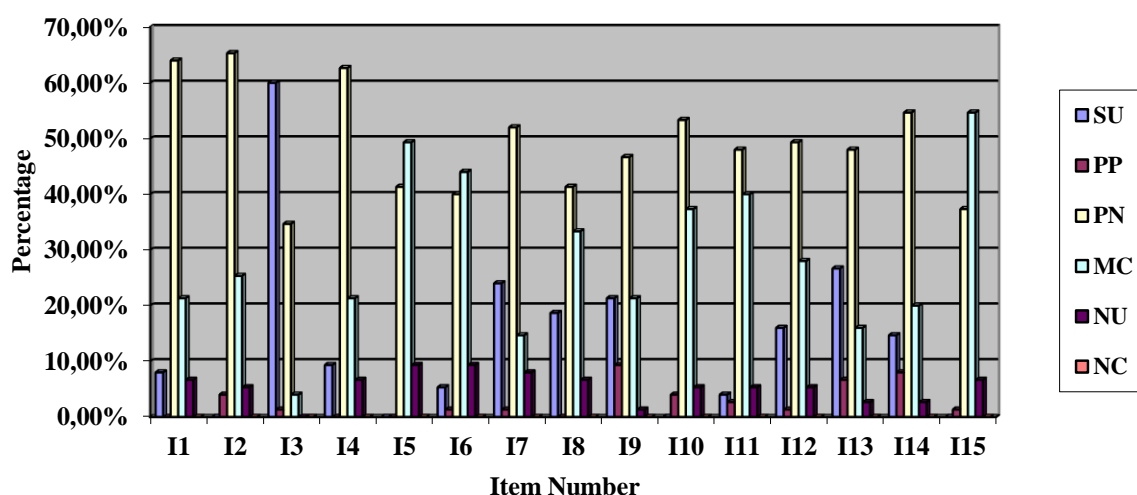
Figure 6. The Conception Categories for Each Item.

The results of the conception category on each item indicate that the highest sound understanding category is item number 3 (60.00%), and the lowest is item numbers 2, 5, 10, and 15 (0.00%). The highest partial positive category is item 9 (9.33%), and the lowest is items 1, 4, 5, and 8 (0.00%). The highest partial negative category is number 2 (65.33%), and the lowest is item 3 (34.67%). The highest misconception category is item number 15 (54.67%), and the lowest is number 3 (4.00%). The category with the highest incidence of "no understanding" is item number 6, accounting for 9.33%, while the lowest is item number 3, registering at 0.00%. Notably, all items exhibit a "no coding" category of 0.00%, indicating that all students responded to all tiers for each item.

The analysis of the four-tier closed-ended format involved the application of the Rasch Model. This analysis aimed to ascertain the validity, reliability, and difficulty level of four-tier closed-ended optical instruments. The outcomes related to instrument validity are presented in Table 6.

Table 6. OUTFIT MNSQ, OUTFIT ZSTD, and PT MEASURE CORR for Each Item of the Four-tier Close-ended Test.

Item Number	Outfit		PT Measure Corr
	MNSQ	ZSTD	
I1	1.46	1.57	0.29
I2	0.62	-1.42	0.40
I3	1.15	0.58	0.51
I4	0.89	-0.32	0.49
I5	0.94	-0.20	0.21
I6	1.39	1.37	0.49
I7	1.14	0.62	0.50
I8	1.42	1.51	0.62
I9	1.31	1.27	0.49
I10	0.66	-1.35	0.57
I11	1.03	0.20	0.48
I12	0.96	-0.07	0.61
I13	1.20	0.89	0.56
I14	1.34	1.30	0.53
I15	0.96	-0.08	0.51

Table 7. The Unidimensionality of Four-tier Closed-ended Test.

Raw variance explained by measures	43.0%
Unexplained variance in 1st contrast (Eigenvalue)	8.1%
Unexplained variance in 1st contrast (Observed)	2.1271

The results of the four-tier closed-ended test analysis presented in Table 6 reveal that items I1 and I5 do not meet the criteria for PT MEASURE CORR. However, they are retained because they satisfy the criteria for OUTFIT MNSQ and OUTFIT ZSTD values. On the other hand, the remaining four-tier test items meet all the criteria for item suitability. The examination of the instrument's unidimensionality evaluates the validity of the Rasch model on each item individually and as a whole. Unidimensionality is a criterion to determine if the developed instrument can effectively measure its intended content. Table 7 illustrates the impact of unidimensionality, showing that the raw variance explained by measures is 43.0%, surpassing the 40% threshold. This result indicates that the overall validity of the four-tier closed-ended test falls into the "good" category (Sumintono & Widhiarsho, 2015). This good category shows that the four-tier closed-ended test has good validity in measuring students' misconceptions. Furthermore, the value of each unexplained variance is less than 15%. As a result, all four-tier closed-ended test items are valid and can be used in total without revision.

Table 8. The Value of Item Reliability, Person Reliability, and Cronbach Alpha.

No	Item	Value
1	Person Reliability	0.65
2	Item Reliability	0.05
3	Cronbach Alpha	0.78

The Rasch Model reliability test consists of Cronbach Alpha values, person reliability, and item reliability. The results of the reliability test utilizing Rasch are shown in Table 8. Based on the results of the four-tiers test analysis in Table 8, item reliability is 0.95, which is categorized as special. Still, personal reliability shows a value of 0.65 and is categorized as weak. Cronbach Alpha has a value of 0.78, which can be categorized as good (Sumintono & Widhiarsho,

2015). The conclusion from the analysis of the four-tier test shows that the items have very good quality, but the consistency of the answers given by students is still weak. In addition, the interaction between students and the questions is good.

Students with conception scores of 35F and 39F have the highest ability, while students with conception scores of 64F have the lowest ability. Although the 35F and 39F students appear to have the best abilities, they still fall short of items

I5 and I15. The 35F and 39F can only answer items I6 and I10 and below. In contrast, 64F students' abilities fall short of all question items. I3 is the item with the lowest level of difficulty. Students 04F, 20F, 63M, 67F, 72F, 14F, 03F, 61M, and 64F have abilities that fall below item I3. As a result, the students struggle to answer item I3 questions correctly. In contrast, I5 and I15 have the highest difficulty levels. There are no students who are capable of answering questions I5 and I15.
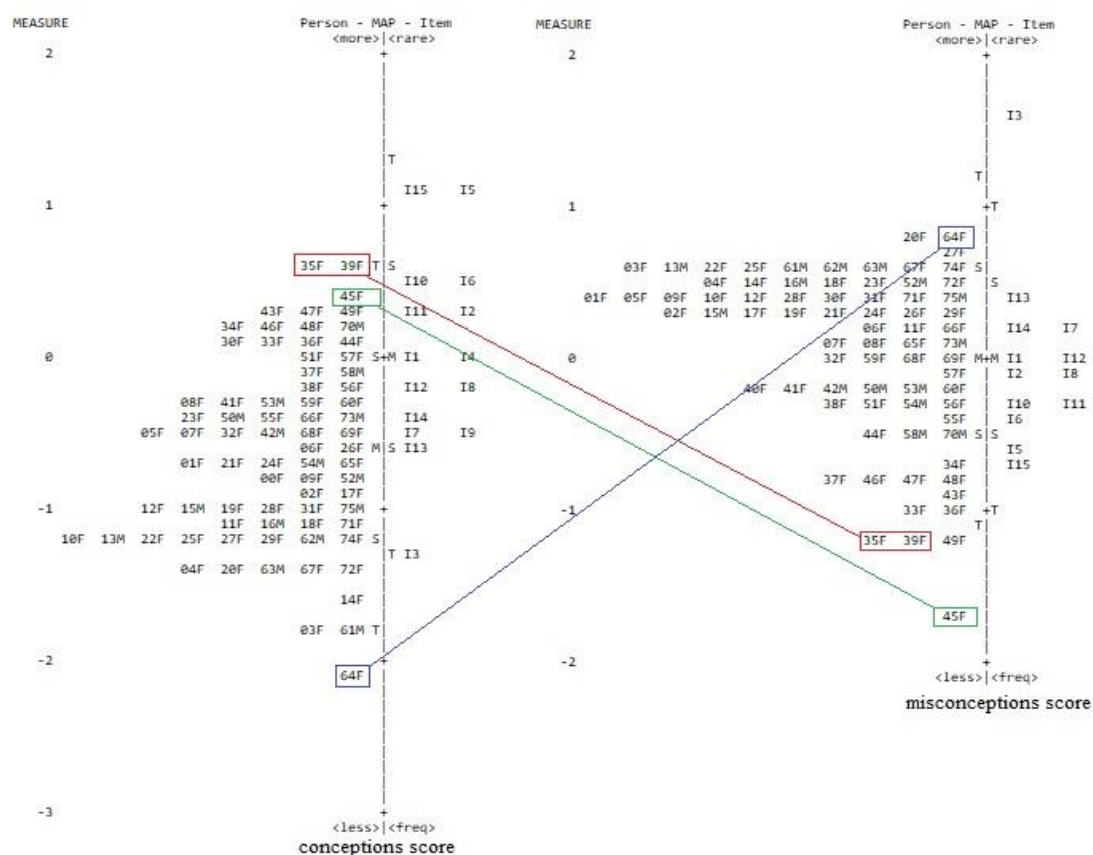
Figure 7. Wright Map Conceptions Score and Wright Map Misconceptions Score.

Based on student misconception scores, students 20F and 64F have the highest misconception scores, while students 45F have the lowest misconception scores. According to Figure 7, all students under I3 indicate that they did not answer with misconceptions. In comparison, item I15 is the item most commonly answered with misconceptions.

The analysis revealed that item I3 had the lowest difficulty in conception, and none of the participants responded with misconceptions. While item I5 presents the most difficult and most frequently answered questions about conception. Students in 64F had the lowest conceptions and the highest misconception scores. Meanwhile, the highest conception value for students of 35F and 39F was not the

student with the lowest misconception. Participants whose lowest misconception scores were 45F, below 35F, and 39F on conception scores. The disparity among students is due to their self-confidence level. This shows that student self-confidence influences student misconceptions. The more students who believe in mistaken concepts, the more students will experience misconceptions.

CONCLUSION

There are four conclusions based on the data analysis and discussion results. First, all items met the CVR scoring criteria, and items I2, I7, I9, I10, I12, I15, and I3 were corrected based on expert advice. Second, students have misunderstandings about each item. Item I5 (49.33 percent) and I15 (49.33 percent) have the most misconceptions (54.67 percent). Third, all items are valid and reliable, with a Cronbach Alpha value of 0.78 in the good category. The fourth conclusion is that conception and misconception are inversely related. The fewer misconceptions, the better the student's understanding, and vice versa. However, misconceptions can also occur due to each student's confidence level. Students with misconceptions about optical instrument materials should be given appropriate treatment, such as appropriate learning. The developed four-level closed test is expected to be used and improved into a better five-level test to investigate the causes of each student's misconceptions.

REFERENCES

Afif, N. F., Nugraha, M. G., & Samsudin, A. (2017). Developing energy and momentum conceptual survey (EMCS) with four-tier diagnostic test items. *AIP Conference Proceedings*, *1848*. https://doi.org/10.1063/1.4983966

Amalia, S. A., Suhendi, E., Kaniawati, I., Samsudin, A., Fratiwi, N. J., Hidayat, S. R., Zulfikar, A., Sholihat, F. N., Jubaedah, D. S., Setyadin, A. H., Purwanto, M. G., Muhaimin, M. H., Bhakti, S. S., & Afif, N. F. (2019). Diagnosis of student's misconception on momentum and impulse trough inquiry learning with computer simulation (ILCS). *Journal of Physics: Conference Series*, *1204*(1). https://doi.org/10.1088/1742-6596/1204/1/012073

Aminudin, A. H., Adimayuda, R., Kaniawati, I., Suhendi, E., Samsudin, A., & Coştu, B. (2019). Rasch analysis of Multitier Open-ended Light-Wave Instrument (MOLWI): Developing and assessing second-years sundanese-scholars alternative conceptions. *Journal for the Education of Gifted Young Scientists*, *7*(3), 557–579. https://doi.org/10.17478/jegys.574524

Anggrayni, S., & Ermawati, F. U. (2019). The validity of Four-Tier's misconception diagnostic test for work and energy concepts. *Journal of Physics: Conference Series*, *1171*(1). https://doi.org/10.1088/1742-6596/1171/1/012037

Bautista, N. U., & Boone, W. J. (2015). Exploring the impact of TeachME™ Lab Virtual Classroom Teaching Simulation on early childhood education majors' self-efficacy beliefs. *Journal of Science Teacher Education*, *26*(3), 237–262. https://doi.org/10.1007/s10972-014-9418-8

Bond, T. G., & Fox, C. M. (2013). Applying the Rasch model fundamental measurement in the human science. In *Applying the Rasch Model*. Routledge. https://doi.org/10.4324/9781410614575

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). Rasch analysis in the human sciences. In *Rasch Analysis in the*

Human Sciences.

Brookhart, S. M., Walsh, J. M., & Zientarski, W. A. (2006). The dynamics of motivation and effort for classroom assessments in middle school science and social studies. *Applied Measurement in Education*, *19*(2), 151–184. https://doi.org/10.1207/s15324818a me1902_5

Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education*, *32*(7), 939–961. https://doi.org/10.1080/09500690902 890130

Coetzee, A., & Imenda, S. N. (2012). Alternative conceptions held by first year physics students at a South African university of technology concerning interference and diffraction of waves. *Research in Higher Education Journal*, *16*(July), 13. https://www.aabri.com/manuscripts/ 121097.pdf

Coştu, B. (2008). Learning science through the PDEODE teaching strategy: Helping students make sense of everyday situations. *Eurasia Journal of Mathematics, Science and Technology Education*, *4*(1), 3–9. https://doi.org/10.12973/ejmste/7530 0

Coştu, B., Ayas, A., & Niaz, M. (2012). Investigating the effectiveness of a POE-based teaching activity on students' understanding of condensation. *Instructional Science*, *40*, 47–67.

Darmana, A., Sutiani, A., Nasution, H. A., Sylvia, N. S. A., Aminah, N., & Utami, T. (2021). Analysis of multi rater with facets on instruments HOTS of solution chemistry based on Tawheed. *Journal of Physics: Conference Series*, *1819*(1).

https://doi.org/10.1088/1742-6596/1819/1/012038

Eckes, T. (2019). *Quantitative data analysis for language assessment volume i chapter Many-facet Rasch measurement Implications for rater-mediated language assessment* (1st Editio). Routledge.

Fariyani, Q., Rusilowati, A., & Sugianto, S. (2017). Four-tier diagnostic test to identify misconceptions in geometrical optics. *Unnes Science Education Journal*, *6*(3), 1724–1729.

Greca, I. M., Seoane, E., & Arriassecq, I. (2014). Epistemological issues concerning computer simulations in science and their implications for science education. *Science and Education*, *23*(4), 897–921. https://doi.org/10.1007/s11191-013-9673-7

Güler, N. (2014). Analysis of open-ended statistics questions with Many Facet Rasch Model. *Eurasian Journal of Educational Research*, *55*, 73–90. https://doi.org/10.14689/ejer.2014.55 .5

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.201 5.1369a

Hammer, D. (1996). More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research. *American Journal of Physics*, *64*(10), 1316–1325. https://doi.org/10.1119/1.18376

Hermita, N., Suhandi, A., Syaodih, E., Samsudin, A., Isjoni, Johan, H., Rosa, F., Setyaningsih, R., Sapriadil, & Safitri, D. (2017). Constructing and

implementing a four tier test about static electricity to diagnose pre-service elementary school teacher' misconceptions. *Journal of Physics: Conference Series*, *895*(1). https://doi.org/10.1088/1742-6596/895/1/012167

Heyd-Metzuyanim, E., & Schwarz, B. B. (2017). Conceptual change within dyadic interactions: The dance of conceptual and material agency. *Instructional Science*, *45*(5), 645–677. https://doi.org/10.1007/s11251-017-9419-z

Husnah, I., Suhandi, A., & Samsudin, A. (2020). Analyzing K-11 students' boiling conceptions with BFT-test using Rasch Model: A case study in the COVID-19 pandemic. *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah*, *5*(2), 225–239. https://doi.org/10.24042/tadris.v5i2.6871

Ismail, I. I., Samsudin, A., Suhendi, E., & Kaniawati, I. (2015). Diagnostik miskonsepsi melalui listrik dinamis four tier test view project conceptual change and mental model in the physics conceptions view project. *Prosiding Simposium Nasional Inovasi dan Pembelajaran Sains 2015*, *June*. https://www.researchgate.net/publication/301523361

Jauhariyah, M. N. R., Suprapto, N., Suliyanah, Admoko, S., Setyarsih, W., Harizah, Z., & Zulfa, I. (2018). The students' misconceptions profile on chapter gas kinetic theory. *Journal of Physics: Conference Series*, *997*(1). https://doi.org/10.1088/1742-6596/997/1/012031

Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. C. (2017). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *Research in Science and Technological Education*, *35*(2), 238–260. https://doi.org/10.1080/02635143.2017.1310094

Kaltakçi, D., & Didiç, N. (2007). Identification of pre-service physics teachers' misconceptions on gravity concept: A study with a 3-tier misconception test. *AIP Conference Proceedings*, *899*(April 2007), 499–500. https://doi.org/10.1063/1.2733255

Kaniawati, I., Rahmadani, S., Fratiwi, N. J., Suyana, I., Danawan, A., Samsudin, A., & Suhendi, E. (2020). An analysis of students' misconceptions about the implementation of active learning of optics and photonics approach assisted by computer simulation. *International Journal of Emerging Technologies in Learning*, *15*(9), 76–93. https://doi.org/10.3991/ijet.v15i09.12217

Koçak, D. (2020). Investigation of rater tendencies and reliability in different assessment methods with many facet Rasch model. *International Electronic Journal of Elementary Education*, *12*(4), 349–358.

Kocakulah, M. S., & Kural, M. (2010). Investigation of conceptual change about double-slit interference in secondary school physics. *International Journal of Environmental and Science Education*, *5*(4), 435–460.

Kucukozer, M. (2010). Peasant rebellions in the age of globalization: The EZLN in Mexico and the PKK in Turkey. *City University of New York.*

Liu, G., & Fang, N. (2016). Student misconceptions about force and acceleration in physics and engineering mechanics education. *International Journal of Engineering Education*, *32*(1), 19–29.

Munawaroh, R., Setyarsih, W., Fisika, J., Matematika, F., Ilmu, D., & Alam, P.

(2016). Identifikasi miskonsepsi siswa dan penyebabnya pada materi alat optik menggunakan three-tier multiple choice diagnostic test. *Jurnal Inovasi Pendidikan Fisika (JIPF)*, *05*(02), 79–81. http://etd.lib.metu.edu.tr/upload/12606

Oberoi, M. (2017). Review of literature on student's misconceptions in science. *International Journal of Scientific Research and Education*, *5(3)*, 6274–6280.

Ohlsson, S., & Cosejo, D. G. (2014). What can be learned from a laboratory model of conceptual change? Descriptive findings and methodological issues. *Science and Education*, *23*(7), 1485–1504. https://doi.org/10.1007/s11191-013-9658-6

Pertiwi, C. A., & Setyarsih, W. (2015). Konsepsi siswa tentang pengaruh gaya pada gerak benda menggunakan instrumen force concept inventory (FCI) Termodifikasi. *Jurnal Inovasi Pendidikan Fisika (JIPF)*, *04*(02), 162–168.

Rohmanasari, F., & Ermawati, F. U. (2020). Using four-tier multiple choice diagnostic test to identify misconception profile of 12th grade students in optical instrument concepts. *Journal of Physics: Conference Series*, *1491*(1). https://doi.org/10.1088/1742-6596/1491/1/012011

Rosita, I., Liliawati, W., & Samsudin, A. (2020). Pengembangan instrumen five-tier newton's laws test (5TNLT) untuk mengidentifikasi miskonsepsi dan penyebab miskonsepsi siswa. *Jurnal Pendidikan Fisika Dan Teknologi*, *6*(2), 297–306. https://doi.org/10.29303/jpft.v6i2.2018

Salamah, L. M., Yulianti, E., & Arif, H. (2017). Pengembangan instrumen diagnostik three-tier untuk mengidentifikasi miskonsepsi siswa SMP pada konsep cahaya. In *Prosiding Seminar Nasional Pembelajaran Ipa Ke-2*. http://ipa.fmipa.um.ac.id/

Salmadhia, F., Rusnayati, H., & Liliawati, W. (2021). Five-tier geometrical optics test feasibility to identify misconception and the causes in high school students. *Berkala Ilmiah Pendidikan Fisika*, *9*(2), 141. https://doi.org/10.20527/bipf.v9i2.8874

Shen, J., Liu, O. L., & Chang, H. Y. (2017). Assessing students' deep conceptual understanding in physical sciences: an example on sinking and floating. *International Journal of Science and Mathematics Education*, *15*(1), 57–70. https://doi.org/10.1007/s10763-015-9680-z

Stein, M., Larrabee, T. G., & Barman, C. R. (2008). A study of common beliefs and misconceptions in physical science. *Journal of Elementary Science Education*, *20*(2), 1–11. https://doi.org/10.1007/bf03173666

Sumintono, B., & Widhiarsho, W. (2015). *Aplikasi pemodelan rasch pada assesment pendidikan*. Trim Komunikata.

Suprapto, N. (2020). Do we experience misconceptions?: An ontological review of misconceptions in science. *Studies in Philosophy of Science and Education*, *1*(2), 50–55. https://doi.org/10.46627/sipose.v1i2.24

Thiagarajan, S. A. O. (1974). *Instructional development for training teachers of exceptional children: a sourcebook*. Indiana Univ., Bloomington. Center for Innovation in Teaching the Handicapped.

Umar, F. A., Samsudin, A., Ramalis, T. R., Sa'diyah, L. H., Dalila, A. A., & Komalasari, K. (2021). Uyo and nanu misconception investigation

(UNAMI) on sound-light waves materials in North Sulawesi. *Journal of Physics: Conference Series*, *2098*(1). https://doi.org/10.1088/1742-6596/2098/1/012027

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*, *45*(3), 197–210. https://doi.org/10.1177/07481756124 40286

Zhou, S., Wang, Y., & Zhang, C. (2016). Pre-service science teachers' PCK: Inconsistency of pre-service teachers' predictions and student learning difficulties in Newton's third law. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(3), 373–385. https://doi.org/10.12973/eurasia.201 6.1203a