**Journal of Education and Learning (EduLearn)**
Vol. 19, No. 2, May 2025, pp. 1144~1156
ISSN: 2089-9823 DOI: 10.11591/edulearn.v19i2.21478 

# A two-tier multiple-choice diagnostic test to find student misconceptions about the change of matter

**Rita Arfi Astuti Ningroom[1], Sri Yamtinah[2], Riyadi[1]**
[1]Department of Elementary School Teacher Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Surakarta, Indonesia
[2]Department of Chemistry Education, Faculty of Teacher Training and Education, Universitas Sebelas Maret, Surakarta, Indonesia

**Article Info**

***Article history:***

Received Nov 11, 2023
Revised May 21, 2024
Accepted Jul 3, 2024

***Keywords:***

Certainty of response index
Change of matter
Diagnostic tests
Misconceptions
Science learning

**ABSTRACT**

There are a lot of very interesting scientific concepts to learn in natural and social science. The initial concepts that the student possesses may contradict the actual concepts, which is what causes misconceptions. Misconceptions are identified using misconception detection test tools. In fact, the development of the use of diagnostic test instruments in Indonesia is still very limited. The objective of the study is to create a diagnostic exam instrument that is two tiers and multiple-choice using certainty of response index (CRI) to identify misunderstandings that students have about changes in matters form in science learning. The research design was research and development (RnD). Development starts with literature studies, design and drafting product drafts, testing product validation with expert validation, and empirical testing. The result analysis showed that the two-tier multiple-choice diagnostic test has a high validity of 0.791. The difficulty level belongs to the moderate category, with a coefficient between 0.2 and 0.8 and a good differential and decaying power. Diagnostic tests are said to be suitable for the detection of misconceptions in science learning in elementary school. The outcome of this study contributes to increasing students' knowledge, understanding, critical thinking, and reducing the level of student misconceptions.



***Corresponding Author:***

Rita Arfi Astuti Ningroom
Department of Elementary School Teacher Education, Faculty of Teacher Training and Education
Universitas Sebelas Maret
St. Kentingan, Ir. Sutami 36, Surakarta, Indonesia
Email: arfirita_astngr@stundent.uns.ac.id

## 1. INTRODUCTION

Education in Indonesia has improved since the introduction of the merdeka curriculum in early 2022. The merdeka curriculum is a form of perfection of the prototype curricula applied to the *Sekolah Penggerak* (driving schools) in Indonesia. The learning activities on the merdeka curriculum are expected to improve the accessibility of learning outcomes of students in a holistic way from both cognitive, affective, and psychomotor competence aspects. Educators should pay attention to the planning and implementation of learning so that no conceptual errors occur in the pupils. One of the topics in the merdeka curriculum at the elementary school is natural and social sciences. Learning natural and social sciences at elementary school makes an important contribution to students in terms of developing scientific thinking skills to solve problems in everyday life [1]. Students who study science are highly motivated to take an active role in protecting the environment and honoring nature and everything in it as a gift from God [2].

Pupils already possess the expertise, experience, and information that make up the foundational ideas in natural science learning before attending school learning activities [3]. The initial concepts of the student may be contrary to the concepts put forward by the experts, which is the cause of misconception. Allen revealed some misconceptions in the natural sciences, one of which was about changing the essence of matter [4]. For example, the loss of fluid in the evaporation process, the process of cloud formation, or the difference in the weight of ice and water. Misunderstandings of concepts experienced by students will cause difficulties in understanding other science concepts and result in poor learning performance [5]. Incorrect or incomplete knowledge stems from the experience of the student, the teacher conveys incorrect information, and misunderstandings in examining the information in the textbook will affect the concept of the learner [3], [6], [7].

Students encounter five different kinds of misconceptions: i) prejudice, ii) non-scientific views, iii) conceptual misunderstandings, iv) regional misconception (vernacular misconception), and v) fact misunderstanding [8]. Correction of conceptual misunderstandings that occur in students is critical to learning achievement [9]. Diagnostic tests become one way to identify misunderstandings among pupils [10], [11]. There are many types of diagnostic tests, one of which is a two-tier multiple-choice diagnostic test. Multi-tier testing is the most widely used diagnostic tool to identify scientific misunderstandings [3]. Using the multi-layer multi-choice exam, which 33.06% of scientific education researchers employed between 2015 and 2019, as the most often utilized evaluation technique [12]. One kind of multitier tool for identifying student misconceptions is the two-stage diagnostic test, which has two tiers that evaluate the material of the scale and the consideration of the student [13]. If students properly respond to the content questions and arguments, they are deemed to have an understanding of science topics. A two-level, multiple-choice exam was created using the relationship between the student's conception and contemplation as the foundation [14].

The development of the use of diagnostic evaluation instruments in Indonesia is still very limited. As with researchers [15], [16], testing instruments are still rarely developed to measure 21st-century skills, and Indonesia is still very open in terms of developing assessment instruments for problem-solving. Therefore, it is necessary to develop a test to determine to what extent the student understands the concept and to measure student misconceptions [17]. A variety of diagnostic tools can be utilized, among others, interviews, open questions, two-level double-optional test instruments, and concept maps [12], [18]. The two-level double-choice test has two levels of choice of answers and reason of choice [19]. However, the results of previous research have not been able to identify students' misconceptions by knowing explicitly pupils' degree of confidence or assurance in completing every single question. The certainty of response index (CRI) is used in this study to create a two-tier multiple-choice exam. CRI participation is used to measure the degree of confidence in student's choice of answers so that teachers can know the level of understanding of students in greater depth [20], [21]. CRI techniques are assessed as effective in identifying misunderstandings in students [22]. This method can identify misunderstandings and pupils who comprehend and don't know and are easy to develop [20].

This study is crucial in assisting educators in determining students' comprehension of IPA concepts, weaknesses as well as strengths of pupils. So, in terms of evaluation, the teacher not only sees the scores obtained by the pupils but also knows the conceptual misplacement experienced by the students. Therefore, a two-tier multiple-choice diagnostic test with CRI is an effective alternative in detecting student misconceptions on sub-explanations of changes like objects. Tes two-tier multiple-choice can assess pupils' comprehension of an idea as it makes it easier for them to make the connections between concepts [23]. Aside from that, what makes this research new is the test instrument's CRI capability, which allows it to distinguish between pupils who lack conceptual understanding and those who have misunderstandings [24].

The goal of this research is to create a diagnostic exam instrument that is two tiers and multiple-choice using CRI to identify misunderstandings that students have about change matters form in learning science. The study's contribution be useful to lessen the misunderstandings that students have about learning. In addition, it helps enhance pupils' comprehension and critical thinking skills.

## 2. METHOD

The design of this research was research and development (RnD). The steps in this research include studying previous research findings, determining product characteristics, developing products, and conducting product trials. The product development cycle in this study was repeated until the product developed is fit for use [25]. The stage in this research consists of the preliminary stage, the development stage, and the testing stage. The subjects of this research were 54 students from State Elementary School of Jeron, State Elementary School of Purworejo, State Elementary School of Sembungan 1 Boyolali Regency, Central Java, Indonesia. The determination of this sample was based on considerations of research subjects experiencing misconception problems in science learning. The data collection techniques used were a test, questionnaires, and an interview The results of the two-level multiple choice instrument test were validated

by experts in the material of science, language, and learning evaluation. The experts who validated this test instrument consisted of 5 expert lecturers and 3 teachers. The validity technique used Aiken's formula with five criteria and a validity coefficient of 0 to 1.00 [26]. Aiken's formula is as (1):

$$V = \Sigma s/[n(c-1)] \quad (1)$$

Descriptions:
s: r–ℓo
c: highest validity assessment number
ℓo: lowest validity assessment number
n: the total number of appraisers

The value categories given by validators consist of five categories. These categories range from irrelevant to very relevant. The following more detailed assessment categories given by validators in validating instrument items are presented in Table 1 [27].

Table 1. Aiken's validity criteria

| Criterion | Value |
|---|---|
| Irrelevant | 1 |
| Less relevant | 2 |
| Quite relevant | 3 |
| Relevant | 4 |
| Very relevant | 5 |

The reliability of the two-tier multiple-choice test instrument in this study used Cronbach's alpha formula. Test instrument items are said to be reliable if they are equal to or exceed 0.7. The CRI identification method in this research aims to detect the occurrence of misconceptions, as well as be able to distinguish them from not understand the concept [28]. The CRI depends on a scale that measures the respondent's trust in responding each question. The level of trust or confidence in the answer is reflected in the CRI scale. Respondents who experience misconceptions can be differentiated simply by comparing whether the answer to a question is correct or not with the level of confidence scale when answering [28]. The following CRI categories in assessing students' level of confidence in answers can be seen in Table 2 [29].

Table 2. Respondent answer criteria as shown by CRI value

| CRI | Criteria | Descriptions |
|---|---|---|
| 0 | Guessing | If the answer to the question is guessed |
| 1 | Uncertain | If you are not sure when answering a question |
| 2 | Confident | If you answer the question confidently |
| 3 | Very confident | If you answer the question very confidently |

## 3. RESULTS AND DISCUSSION

The RnD that has been carried out begins with a preliminary study. The preliminary study aims to find out the problems and necessary needs. The preliminary study carried out included interviews and distributing questionnaires. Based on researchers' findings regarding teachers' needs in the field, it is known that teachers only use formative and summative test instruments in learning activities. Even though the assessment instruments used by teachers are in the form of ordinary multiple-choice or short entries or descriptions, they are unable to differentiate between students who understand the concept and have misconceptions and those who do not understand the concept. As the class teacher said:

> … "*Yes, usually only formative tests (daily tests) and summative tests (mid-semester assessment or final semester assessment). I have never used diagnostic tests in science learning*" …

Most classroom teachers have never developed diagnostic test instruments. Teachers thought that diagnostic tests were not very important and did not know the purpose of carrying out diagnostic tests, either at the beginning of learning or at the end of learning. Assessment instruments that should be developed by the provisions for implementing the Independent Curriculum are diagnostic assessment instruments, formative assessment, and summative assessment.

> … "*At the end of the sub-chapter discussing material and changes, I created questions for daily assessment. Then, near the end of the semester, I made summative questions. The questions I make are usually multiple-choice, short answers, and descriptions. Sometimes I just copy questions that are already in the book*" …

Class teachers realize that the instruments currently available do not meet learning needs. Instructors acknowledged that they were ignorant about the needs, shortcomings, and learning challenges of their pupils. Lastly, teachers are unable to distinguish between pupils who get the topic already, those who have misunderstandings, and those who do not.

> … "*Usually, I ask students. Have they understood the material or not? I have never created a test to detect misconceptions in them. The assessment I carry out is only to determine students' learning achievements (completed or not)*" …

The results of interviews with class teachers were supported by the results of a needs analysis questionnaire that was given to 54 students. The aim of distributing the questionnaire is to gather information about science learning activities in class and to find out the needs of students, especially in assessment activities. As many as 82.15% of students still have difficulty understanding the topic of material and changes in the states of substances in science learning. The majority of students experienced differences in knowledge with existing knowledge regarding the topic of changes in the state of substances, reaching 86%. As many as 85.57% of students had never been assessed at the start of learning (diagnostic assessment) and 72.3% had difficulty working on the assessment questions given by the teacher.

Based on the preliminary studies that have been carried out, researchers are trying to develop diagnostic test instruments to detect misconceptions in students. The product development stage begins with preparing a product draft. The procedure for preparing a product draft is to determine the purpose of the test, arrange the test grid, determine the form of the test, and determine the length of the test. Next, the draft two-tier multiple-choice diagnostic test instrument product was tested to determine its feasibility. Product trials are divided into three stages, namely limited-scale trials, medium-scale trials, and wide-scale trials. Limited trials were carried out in three schools with five students in each school. Limited trials were carried out by reading instruments and distributing questionnaires. Meanwhile, medium-scale and wide-scale trials aim to see the quality of the instrument in terms of constructs including validity, reliability, level of difficulty, and distinguishability. Meanwhile, extensive testing was carried out to see the effectiveness of the test instruments in detecting misunderstandings of concepts experienced by students.

A test instrument is said to be suitable for use if the test instrument goes through several stages of validation. The validation carried out is in the form of content validation and construct validation. Content validation was carried out by five validators from lecturers and three experienced elementary school teachers. Meanwhile, construct validity is carried out by calculating the validity, reliability, level of difficulty, and differentiation of each item.

### 3.1. Question quality analysi

Aiken's content validity was calculated using the results of the assessment of a two-tier multiple-choice diagnostic test instrument from eight assessors who work as expert lecturers and senior teachers. Analysis of fifteen question items using the Microsoft Excel program shows a validity value of 0.791 to 1, which means that each question item has high validity. Briefly, the validity analysis of the question items is presented in Table 3.

Table 3. Aiken's validity using Microsoft Excel

| Item | i-1 | i-2 | i-3 | i-4 | i-5 | i-6 | i-7 | i-8 | i-9 | i-10 | i-11 | i-12 | i-13 | i-14 | i-15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| V. Aiken's | 0.958 | 0.916 | 0.916 | 0.875 | 0.916 | 0.958 | 0.791 | 0.916 | 0.875 | 1 | 0.958 | 0.833 | 0.916 | 0.916 | 0.916 |

The first stage of testing is a limited-scale trial. This testing was carried out with the help of students who were randomly selected in each representative school. The two-tier multiple-choice diagnostic test questions were distributed to fifteen students to read. After that, a questionnaire was given which aimed to find out responses and input from students regarding the readability of each item on the instrument. The results of limited-scale testing are presented in Table 4.

Table 4. Limited scale trial results

| No | Question | Answer (%) | |
|---|---|---|---|
| | | Yes | No |
| 1 | Are the question sentences clear and easy to understand? | 33.33 | 66.67 |
| 2 | Is the writing in these letters easy to read? | 86.67 | 13.33 |
| 3 | Is the language in the questions easy to understand? | 40 | 60 |
| 4 | Are the images presented clear and easy to understand? | 93.33 | 6.67 |
| 5 | Are the answer choices varied/diverse? | 93.33 | 6.67 |
| 6 | Are the reasons for the answer choices easy to understand? | 46.67 | 53.33 |
| 7 | Are the answer choices and reason choices appropriate? | 80 | 20 |
| 8 | Is there enough time to complete the questions? | 93.33 | 6.67 |

Based on the results of limited trials in Table 4, it is known that as many as 66.67% of students rated the question sentences as unclear and difficult to understand, 60% of students stated that the meaning of the language in the question items was difficult to understand and the same was true for 53 choices of reasons for answers. 33%. The two-tier multiple-choice diagnostic test instrument in the initial product given to students still needs improvement, especially in terms of grammar, both sentences, and the language used. Based on the results of the questionnaire, it is necessary to improve the question items which are difficult to understand in sentences or language. An alternative that can be done is to change the sentence or add an image so that students better understand the meaning of the question. Meanwhile, the image aspects, variations in answer choices, letters used, suitability of answers and reasons, and processing time are sufficient. The following are the results of the revised test question items in the limited trial and are presented in Table 5.

Table 5. Results of revision of question items in limited trials

| No. Question | Initial form of questions | Improvement results |
|---|---|---|
| 1 | Books and cupboards are examples of objects….<br>a. Liquid<br>b. Congested<br>c. Gas<br>d. Liquid | Bricks are an example of an object that has the shape of...<br>a. Liquid<br>b. Congested<br>c. Gas<br>d. Box |
| 6 | Objects that always have a constant shape are objects...<br>a. Gas<br>b. Liquid<br>c. Congested<br>d. Hard | Tini took the water from the bottle and then put the water into the glass. The following statement is appropriate to the activities carried out by Tini, namely...<br>a. The volume of water changes<br>b. The shape of the water remains<br>c. The water changes according to the container<br>d. The amount of water increases |
| 14 | An example of a melting event is...<br>a. Make jelly<br>b. Make ice lollies<br>c. Blowing glass<br>d. Melting ice cream | The change from solid to liquid occurs when...<br>a. Make pudding<br>b. Make ice lollies<br>c. Blowing glass<br>d. Heat margarine |

The second stage of testing is a medium-scale trial. The initial product, which had been improved in the previous testing stage, was given to students in each class at each representative school. A total of 54 students were given the same questions to work on. After completing the questions, the student's test results are analyzed to determine the quality of the questions which include reliability, discrimination, level of difficulty, and distraction index. Question quality analysis was carried out using ITEMAN software. The results of the analysis at this testing stage are used as a basis for revising the product to make it more suitable for use as an assessment instrument. The third stage of testing is the final test of the test instrument. The third stage of testing is a wide-scale trial. The fourth stage testing mechanism is more or less the same as the third stage testing. The difference lies in the number of respondents who take the test questions. The following are the results of the analysis of differentiating power, level of difficulty, and distraction index in medium-scale trials in Table 6.

Table 6 shows that several questions have a distractor index value of 0.000. This means that the instrument was not chosen at all by the test takers and must be corrected. Apart from that, there are still several questions that have negative differentiating power values. This means that students in the high ability category tend to choose the wrong answer, while students with low ability tend to answer questions correctly. Improvements to question item number 1 were made to tier 1 answer choice (distractor) D because this option was not yet effective. Improvements to question item number 4 were made to answer choices

(distractors) A and D in tier 1. Meanwhile, improvements to question item number 10 were made to answer choice C and improvements to the use of question sentences so that students could more easily understand the meaning of the question. Meanwhile, improvements to question number 11 include tier 1 answer options B and C (distractor). Improvements were also made at tier 2 on items with negative differentiating power values, namely numbers 1, 4, 10, and 11.

Table 6. Results of analysis of discriminating power, level of difficulty, and distraction index in medium-scale trials

| No | Tier | Differentiating power | Difficulty level | Distraction index | | | |
|---|---|---|---|---|---|---|---|
| | | | | A | B | C | D |
| 1 | Tier 1 | 0.331 | 0.833 | | | | 0.000 |
| | Tier 2 | -0.078 | 0.760 | | | | |
| 4 | Tier 1 | 0.267 | 0.667 | 0.000 | | | 0.000 |
| | Tier 2 | -0.061 | 0.667 | | | | |
| 10 | Tier 1 | 0.460 | 0.250 | | | 0.000 | |
| | Tier 2 | -0.021 | 0.667 | | | | |
| 11 | Tier 1 | 0.577 | 0.750 | | 0.000 | 0.000 | |
| | Tier 2 | -0.067 | 0.750 | | | | |

The third trial is the field implementation trial (wide scale) and final revision (operational field test and final product revision). At this stage, the two-tier multiple-choice instrument equipped with CRI was given to 6 classes in different schools, with high, medium, and low categories. The instrument is given back to students to determine the performance of the instrument after improvements have been made in medium-scale trials. The following results of the analysis of discriminating power, level of difficulty, and distraction index in wide-scale trials are presented in Table 7.

Table 7. Results of analysis of discriminating power, level of difficulty, and distraction index in wide-scale trials

| No | Tier | Differentiating power | Difficulty level | Distraction index | | | |
|---|---|---|---|---|---|---|---|
| | | | | A | B | C | D |
| 1 | Tier 1 | 0.556 | 0.698 | | 0.210 | 0.230 | |
| | Tier 2 | 0.538 | 0.484 | | | | |
| 4 | Tier 1 | 0.325 | 0.516 | 0.080 | | | 0.050 |
| | Tier 2 | 0.601 | 0.560 | | | | |
| 7 | Tier 1 | 0.626 | 0.308 | 0.030 | | 0.070 | 0.100 |
| | Tier 2 | 0.616 | 0.473 | | | | |
| 10 | Tier 1 | 0.575 | 0.505 | | | 0.210 | |
| | Tier 2 | 0.381 | 0.368 | | | | |
| 12 | Tier 1 | 0.494 | 0.588 | | | 0.045 | |
| | Tier 2 | 0.491 | 0.341 | | | | |

Table 7 shows that the five questions that have been corrected have a good distractor index value, namely above 0.02. This means that the answer choices and reason choices for each test item function well. The discriminating power index is positive, meaning that students in the high-ability category tend to choose the correct answer, while students with low ability tend to choose the wrong answer or just guess. The difficulty level of the questions is included in the medium category, so it can be said that the test questions can be used. Table 8 shows the reliability of the two-tier multiple-choice diagnostic test instrument.

Table 8. Reliability of the two-tier multiple choice (TTMC) diagnostic test instrument

| Test phase | Tier | Reliability value | Conclusion |
|---|---|---|---|
| Limited scale trials | - | - | - |
| Medium scale trial | Tier 1 | 0.765 | Reliable |
| | Tier 2 | 0.729 | Reliable |
| Wide-scale trials | Tier 1 | 0.783 | Reliable |
| | Tier 2 | 0.741 | Reliable |

An instrument is said to be reliable if its reliability coefficient is more than equal to 0.70 (ri1≥0.70) [30], [31]. The reliability of the two-tier multiple-choice diagnostic test instrument in medium-scale trials was 0.765 in the first tier and 0.729 in the second tier. Meanwhile, in the wide-scale trial, the test instrument

reliability results were 0.783 in the first tier and 0.741 in the second tier. Assessment must follow the characteristics of a good test or assessment, namely: i) valid, ii) reliable, iii) objective, and iv) practical so that it can be used well [32]. This is the case with the results of the empirical analysis of the diagnostic test instruments in this research. The items in this diagnostic test instrument have met the criteria for internal consistency and high reliability for Cronbach's alpha (0.82) [33]. Thus, it can be concluded that the two-tier multiple-choice diagnostic test instrument developed meets the reliable criteria.

Based on Table 9, a percentage comparison of the differentiating power in each category can be obtained, namely bad, fair, good, and very good. In the medium scale trial, the poor category in tier 2 had a percentage of 26.67%, the fair category in tier 1 was 40% and tier 2 was 20%, the good category in tier 1 was 13.33% and tier 2 was 26.67%, as well as the very good category in tier 1 at 46.67% and tier 2 at 26.67%. Meanwhile, in the wide-scale trial, questions in the good category in tier 1 amounted to 13.33% tier 2 amounted to 6.67%, and the very good category in tier 1 amounted to 86.67% and tier 2 amounted to 93.33%. The percentage comparison of the differentiating power of the questions can be seen in Figure 1.

Table 9. Differentiating power of the TTMC diagnostic test instrument

| No. Question | Tier | Differentiating power | | | |
|---|---|---|---|---|---|
| | | Medium scale trial | | Wide scale trial | |
| | | Point biser (%) | Description | Point biser (%) | Description |
| 1 | Tier 1 | 0.331 | Good | 0.556 | Very good |
| | Tier 2 | -0.078 | Poor | 0.538 | Very good |
| 2 | Tier 1 | 0.440 | Very good | 0.475 | Very good |
| | Tier 2 | 0.369 | Good | 0.594 | Very good |
| 3 | Tier 1 | 0.285 | Fair | 0.335 | Good |
| | Tier 2 | 0.391 | Good | 0.536 | Very good |
| 4 | Tier 1 | 0.267 | Fair | 0.325 | Good |
| | Tier 2 | -0.061 | Poor | 0.601 | Very good |
| 5 | Tier 1 | 0.500 | Very good | 0.649 | Very good |
| | Tier 2 | 0.375 | Good | 0.629 | Very good |
| 6 | Tier 1 | 0.500 | Very good | 0.664 | Very good |
| | Tier 2 | 0.421 | Very good | 0.604 | Very good |
| 7 | Tier 1 | 0.110 | Fair | 0.626 | Very good |
| | Tier 2 | 0.078 | Fair | 0.616 | Very good |
| 8 | Tier 1 | 0.258 | Fair | 0.438 | Very good |
| | Tier 2 | 0.229 | Fair | 0.648 | Very good |
| 9 | Tier 1 | 0.376 | Good | 0.426 | Very good |
| | Tier 2 | 0.401 | Very good | 0.412 | Very good |
| 10 | Tier 1 | 0.460 | Very good | 0.575 | Very good |
| | Tier 2 | -0.021 | Poor | 0.381 | Good |
| 11 | Tier 1 | 0.201 | Fair | 0.591 | Very good |
| | Tier 2 | 0.310 | Good | 0.525 | Very good |
| 12 | Tier 1 | 0.577 | Very good | 0.494 | Very good |
| | Tier 2 | -0.067 | Poor | 0.491 | Very good |
| 13 | Tier 1 | 0.423 | Very good | 0.680 | Very good |
| | Tier 2 | 0.296 | Fair | 0.513 | Very good |
| 14 | Tier 1 | 0.463 | Very good | 0.644 | Very good |
| | Tier 2 | 0.517 | Very good | 0.639 | Very good |
| 15 | Tier 1 | 0.277 | Fair | 0.529 | Very good |
| | Tier 2 | 0.409 | Very good | 0.591 | Very good |

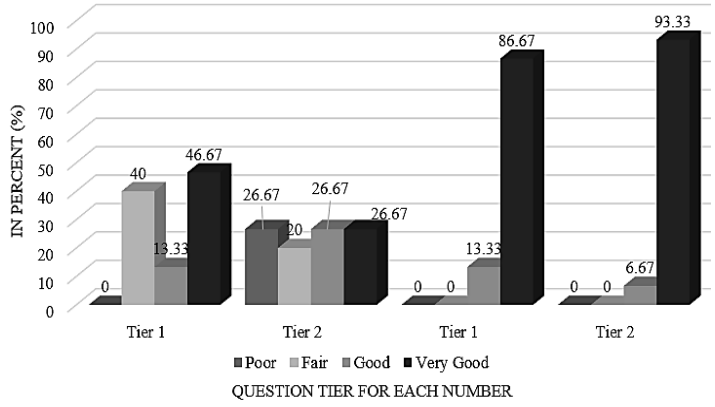


Figure 1. Percentage of differentiating power in medium scale and wide scale trials

The level of difficulty states the proportion of test takers who answered correctly on each question item to all test takers. The level of difficulty is symbolized by the letter P, which means proportion. The greater the P value, the easier the question's difficulty level, and the smaller the P value, the more difficult the question's difficulty level. One of the criteria for a good test question is that it has a difficulty level that is not too difficult or not too easy. Based on the results of the analysis of the questions using ITEMAN software, the level of difficulty can be determined from the value obtained in the prop correct column. The following is presented regarding the difficulty level of the two-tier multiple-choice diagnostic test instrument in Table 10.

Table 10. Difficulty level of the two-tier multiple-choice diagnostic test instrument

| No. Question | Tier | Difficulty level | | | |
|---|---|---|---|---|---|
| | | Medium scale trial | | Wide scale trial | |
| | | Prop.correct (%) | Description | Prop.correct (%) | Description |
| 1 | Tier 1 | 0.833 | Easy | 0.698 | Middle |
| | Tier 2 | 0.760 | Easy | 0.484 | Middle |
| 2 | Tier 1 | 0.667 | Middle | 0.401 | Middle |
| | Tier 2 | 0.833 | Easy | 0.445 | Middle |
| 3 | Tier 1 | 0.583 | Middle | 0.412 | Middle |
| | Tier 2 | 0.667 | Middle | 0.500 | Middle |
| 4 | Tier 1 | 0.667 | Middle | 0.516 | Middle |
| | Tier 2 | 0.667 | Middle | 0.560 | Middle |
| 5 | Tier 1 | 0.750 | Easy | 0.621 | Middle |
| | Tier 2 | 0.667 | Middle | 0.582 | Middle |
| 6 | Tier 1 | 0.250 | Difficult | 0.571 | Middle |
| | Tier 2 | 0.583 | Middle | 0.429 | Middle |
| 7 | Tier 1 | 0.667 | Middle | 0.308 | Middle |
| | Tier 2 | 0.750 | Easy | 0.473 | Middle |
| 8 | Tier 1 | 0.833 | Easy | 0.330 | Middle |
| | Tier 2 | 0.916 | Easy | 0.357 | Middle |
| 9 | Tier 1 | 0.750 | Easy | 0.654 | Middle |
| | Tier 2 | 0.667 | Middle | 0.396 | Middle |
| 10 | Tier 1 | 0.250 | Difficult | 0.505 | Middle |
| | Tier 2 | 0.667 | Middle | 0.368 | Middle |
| 11 | Tier 1 | 0.916 | Easy | 0.324 | Middle |
| | Tier 2 | 0.760 | Easy | 0.418 | Middle |
| 12 | Tier 1 | 0.750 | Easy | 0.588 | Middle |
| | Tier 2 | 0.750 | Easy | 0.341 | Middle |
| 13 | Tier 1 | 0.583 | Middle | 0.434 | Middle |
| | Tier 2 | 0.916 | Easy | 0.341 | Middle |
| 14 | Tier 1 | 0.833 | Easy | 0.643 | Middle |
| | Tier 2 | 0.833 | Easy | 0.346 | Middle |
| 15 | Tier 1 | 0.750 | Easy | 0.538 | Middle |
| | Tier 2 | 0.667 | Middle | 0.379 | Middle |

The distribution of questions is uneven in medium-scale trials and wide-scale trials. Questions in the easy category in the medium scale trial were (53.33%), the medium category was (33.33%) and the difficult category was (13.34%). Meanwhile, in the wide-scale trial, there were no questions in the easy or difficult categories, all the questions were in the medium category. The percentage comparison of the level of difficulty of the questions can be seen in Figure 2.

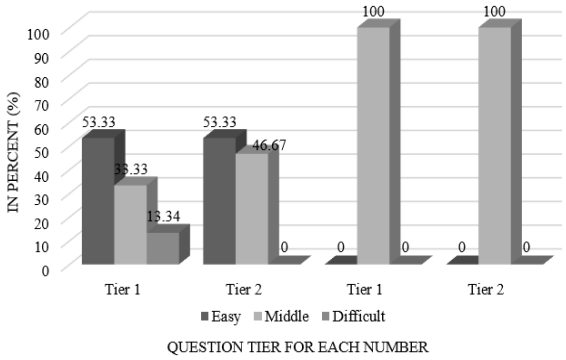


Figure 2. Percentage of difficulty levels on medium scale and wide scale trials

### 3.2. Two-tier multiple-choice diagnostic test instrument to detect student misconceptions

Previous research stated that learning will be effective if teachers can understand students' difficulties and misconceptions. Therefore, it is necessary to carry out diagnostic measures to detect student misconceptions. One way to diagnose misconceptions is to use diagnostic instruments. Diagnostic tests are useful for identifying student difficulties and for planning efforts to solve identified difficulties. The appropriate form of diagnostic test instrument is a two-tier multiple-choice equipped with CRI. Cullinane and Liston [34], stated that the inclusion of reasons in tier 2 may be applied to enhance higher-order thinking abilities (HOTS) and assess pupils' capacity for justification. The two-tier multiple-choice form has weaknesses because it is not always accurate in distinguishing students' levels of understanding. Therefore, to overcome this weakness, the CRI technique was included. The CRI scale is useful for knowing the respondent's level of confidence in answering the first and second tiers.

The importance of using test instruments that can detect students' misconceptions in science learning in the classroom is a strong reason for developing a TTMC instrument equipped with CRI to detect students' misconceptions in science learning which can be seen from the test taker's confidence in the answers given and the consistency of the answers where if the test taker consistently answers correctly then the student is declared to understand the concept if he consistently gives wrong answers or misconceptions then the test taker is declared to have a misconception and likewise if the test taker shows inconsistent answers then it can be stated that they do not understand the concept.

The development of a two-tier multiple-choice test instrument equipped with the CRI. Apart from being able to measure students' cognitive abilities, this instrument is also able to detect students' misconceptions. The development of this instrument is based on indicators of competency achievement. The following are the criteria for the level of understanding of the concept which are depicted in Table 11.

Table 11. Concept understanding level category

| No | Answer | CRI | Reason | CRI | Category |
|---|---|---|---|---|---|
| 1 | Right | High | Right | High | U |
| 2 | Right | High | Right | Low | DU |
| 3 | Right | Low | Right | High | DU |
| 4 | Right | High | False | High | M |
| 5 | False | High | Right | High | M |
| 6 | Right | High | False | Low | DU |
| 7 | Right | Low | False | High | DU |
| 8 | False | Low | Right | Low | DU |
| 9 | False | High | False | High | U |
| 10 | False | Low | False | High | DU |
| 11 | False | High | False | Low | DU |

*Note: U=Understand, DU=Don’t Understand, M=Misconception

Table 11 is a guide to criteria for understanding concepts to identify misconceptions that occur in students. To determine the level of students' understanding of concepts, the author here combines the two-tier model developed by David Treagust and the CRI developed by Saleem Hasan. By combining the two instruments, it is hoped that students' misconceptions regarding the concept of matter (a form of matter) and its changes can be identified. The results of the analysis of students' level of understanding in the field implementation trials are presented in Table 12.

Table 12. Analysis of students' concept understanding level in field implementation trials

| Student understanding level | Result (%) |
|---|---|
| Understanding concepts (U) | 41.88 |
| Not understanding the concept (DU) | 29.91 |
| Misconceptions (M) | 28.21 |

Table 12 shows the results of the analysis of students' level of understanding of concepts at the field implementation trial stage. The percentage of students with a level of understanding of the concept was 41.88%, not understanding the concept was 29.91%, and misconceptions were 28.21%. At this test stage, the percentage of students who have misconceptions and don't comprehend the idea has gone down, while the percentage of students who understand the concept has gone up. In this research, students' misconceptions can be identified by the student's level of confidence in providing answers. If a student has a high level of confidence in this matter with a CRI of 3 to 5, but the answer given is wrong then it is stated that the student has a misconception. If

students have a high level of confidence in this case with CRI at a value of 3 to 5 and the answer is correct, then the student is declared to understand the concept. Meanwhile, if students have a low level of confidence, but the answers given are right or wrong, then the students are said to not understand the concept.

Consequently, it can be said that this study's development of a TTMC test instrument with CRI can lower or lower the percentage of misconceptions and conceptual misunderstandings in the material on object shape changes while raising the percentage of conceptual understanding. The findings of this study corroborate those of other studies that showed TTMC could assess students' conceptual knowledge, misunderstandings, and ignorance of the respiratory system content [35]. The study's findings also show that 37.7% of students understood the concept of the respiratory system content; 41.6% of students had misconceptions; and 20.7% of students didn't grasp the concept. Then, this research is also in line with previous research which states that TTMC which has been developed and tested for content validity, construct, and reliability can be used by practitioners to diagnose students' initial concepts, whether they already understand the concepts or whether there are still misconceptions [36]. Apart from identifying misconceptions, TTMC may be used to assess pupils' HOTS [37]. Furthermore, previous relevant research also demonstrated that the inclusion of CRI to assess students' degree of confidence in their responses and the introduction of TTMC to demonstrate students' level of comprehension and further thought [38]. The CRI provides a gauge for how certain a student is in his or her ability to choose and apply concepts, laws, or information to arrive at the desired response [39].

The development of a two-tier multiple-choice diagnostic test instrument equipped with CRI. Apart from being able to measure students' mastery of concepts, can also differentiate and identify students who do not understand the concepts and students who experience misconceptions. The complete distribution of students' conceptual understanding can be seen in Table 13.

Table 13. Distribution of students' understanding of concepts

| No | Indicators of competence achievement | M (%) | U (%) | DU (%) |
|---|---|---|---|---|
| 1 | Identify the shape or form of objects | 21.67 | 40.00 | 38.33 |
| 2 | Analyzing the form of matter in an object | 21.33 | 36.67 | 42.00 |
| 3 | Explain the change in form | 19.33 | 45.00 | 35.67 |
| 4 | Distinguish between change of form events | 28.00 | 30.00 | 40.33 |
| 5 | Analyzing shape change events | 13.00 | 43.67 | 43.67 |

*Note: U=Understand, DU=Don’t understand, M=Misconception

Table 13 shows that the distribution of students' conceptual understanding is presented in the competency achievement indicator (GPA) data. The results above show that there are students who still have wrong concepts on the sub-discussion topic of material forms and their changes. For example, in question number 11, students do not yet understand the process of changing state from gas to solid, this indicates that students do not understand the concept of changing from one state to another well. Students are still confused between sublimation and crystallization changes. In this case, the type of misconception that occurs is included in the category of preconceived notions. This type of misconception occurs because students' initial conceptual understanding was wrong regarding the concepts of sublimation and crystallization for these indicators. Understanding the initial wrong concept gives rise to misconceptions in students.

However, the results of this research can conclude that the level of understanding of students in understanding the concept of material changes in the form of objects in science learning using TTMC with CRI is greater than students who experience misconceptions. The results of this research correlate with previous research, which revealed that the development of the TTMC test can measure students' misconceptions about science learning [17]. The test instruments developed in previous research were tested for validity and reliability only through trials on students. Meanwhile, before the research test instrument was tested on a limited and extensive basis on students, it had received validity from experts in the fields of learning materials, language, and learning evaluation.

The results of this research show that the development of TTMC equipped with CRI is better at measuring students' misconceptions using the RnD research method. This is corroborated by earlier studies that measure a person's misunderstandings using the CRI, which gauges an individual's answer to questions [40]–[42]. This is different from previous research in developing TTMC using the qualitative descriptive research method [43] although without reducing the benefits in identifying cases of misconceptions among students in learning. The results of this research were developed through the RnD method so that the results of the instrument product followed the results of the initial study of student needs.

Several ways that can be done to reduce and prevent misconceptions in students are knowing about several causes and the ability to diagnose [44]. Apart from that, using the Cocoaer learning model [45]. The two main theories that underpin the Cocoaer learning model are: i) Vygotsky's social constructivism theory,

which emphasizes social learning and scaffolding, particularly for students in their zone of proximal development; and ii) Bandura's social learning theory acknowledges that learning happens by watching the behavior of people and the environment (in which processing information includes attention, memory, creation, and motivation). Cocoaer learning models with the syntax: i) committee and expose beliefs, ii) confront beliefs, iii) assimilate and accommodate the concept, iv) extend the concept, and v) reflect beliefs.

The results of this research contribute to increasing students' knowledge, understanding, critical thinking, and higher thinking. Apart from that, the development of TTMC accompanied by CRI can also reduce the level of student misconceptions. The results of this research are more resilient than previous research because the development of the TTMC test has two levels which function to answer questions on the topic of changes in the state of objects in science learning and the reasons for answering this topic. Apart from that, CRI is useful for knowing students' level of confidence in answering these questions. Future research may explore more widely other learning materials related to science and social learning.

However, this research was only developed within the district scope at the elementary school level so it needs to be developed again with a wider scope and at a higher school level. Apart from that, this TTMC instrument needs to be further developed with other subject matter concepts and can be developed into a three-tier or four-tier multiple-choice development so that students' ability to understand concepts is further improved. This is necessary because it requires further and in-depth study to ensure students' confidence in answering each question on a topic and the reasons for answering that topic.

## 4. CONCLUSION

The purpose of this study is to develop a two-tiered, multiple-choice diagnostic test tool that uses CRI to identify students' misconceptions about change matters form in learning science. Based on the findings of the content validity study, it is possible to infer that the two-tier multiple-choice diagnostic test instrument has high validity, namely 0.791 to 1, according to the expert's score and the calculation of content validity using Aiken's formula. If the validity coefficient exceeds 0.75, the instrument is considered to have excellent validity. Meanwhile, the construct validity study revealed that numerous items did not match the standards, thus they were improved. The diagnostic test instrument's overall difficulty index falls into the medium group, with a coefficient ranging from 0.2 to 0.8. Two items had limited distinguishing power; thus, they were changed to be usable. A total of 45 distractors were used, 5 of which (11.11%) did not work well since less than 5% of respondents picked them. The designed diagnostic test instrument is practicable and may be used to identify primary school students' misunderstandings in scientific classes, as well as content sub-discussion and adjustments. The degree of knowledge of students' concepts in the main field trial stage and field implementation trials was varied. Specifically, the proportion of those who grasped the concept was 41.88%; those who did not comprehend it was 29.91%; and those who had misconceptions were 28.21%. At this test stage, the percentage of students who have misconceptions and don't comprehend the idea has gone down, while the percentage of students who understand the concept has gone up. According to indicators of skill accomplishment, there are still kids who don't comprehend the topic of changing from one form to another. Students are indicated to experience misconceptions of the preconceived notions type, namely initial concepts that are wrong, giving rise to misconceptions in subsequent concepts. This research results and discussion provide a TTMC instrument equipped with CRI to detect misconceptions in students. Therefore, to overcome misconceptions in the future, a learning model is needed to reduce and overcome problems regarding misconceptions.

## ACKNOWLEDGEMENTS

Thank you to the Elementary School Teacher Education, Faculty of Teacher Training and Education at Universitas Sebelas Maret for allowing permission for researchers to do RnD.

## REFERENCES

[1] E. P. Dewi and F. Wulandari, "Identification of misconceptions in science learning during the COVID-19 pandemic using the CRI (certainty of response index) method for primary school students," *Jurnal Penelitian Pendidikan IPA*, vol. 7, no. SpecialIssue, pp. 145–150, Nov. 2021, doi: 10.29303/jppipa.v7iSpecialIssue.876.

[2] N. M. K. Yulianti, I. G. A. O. Negara, and I. W. Sujana, "Contribution of Naturalist intelligence and learning participation toward students' knowledge competence in science," *International Journal of Elementary Education*, vol. 4, no. 3, p. 366, Sep. 2020, doi: 10.23887/ijee.v4i3.26298.

[3] S. Soeharto, "Development of adiagnostic assessment test to evaluate science misconceptions in terms of school grades: a rasch measurement approach," *Turkish Journal of Science Education*, vol. 18, no. 3, pp. 351–370, Sep. 2021, doi: 10.36681/tused.2021.78.

[4] M. Allen, *Misconceptions in primary science*. Unity Kingdom: McGraw-Hill Education, 2014.

[5] S. Soeharto and B. Csapó, "Exploring Indonesian student misconceptions in science concepts," *Heliyon*, vol. 8, no. 9, p. e10720, Sep. 2022, doi: 10.1016/j.heliyon.2022.e10720.

[6] A. Hakim, Liliasari, A. Kadarohman, and Y. M. Syah, "Effects of the natural product mini project laboratory on the students conceptual understanding," *Journal of Turkish Science Education*, vol. 13, no. 2, pp. 27–36, 2016.

[7] Z. D. Kirbulut and O. Geban, "Using three-tier diagnostic test to assess students' misconceptions of states of matter," *EURASIA Journal of Mathematics, Science and Technology Education*, vol. 10, no. 5, pp. 509–521, Dec. 2014, doi: 10.12973/eurasia.2014.1128a.

[8] N. Suprapto, "Do we experience misconceptions?: an ontological review of misconceptions in science," *Studies in Philosophy of Science and Education*, vol. 1, no. 2, pp. 50–55, Apr. 2020, doi: 10.46627/sipose.v1i2.24.

[9] F. O. Pribadi, Supahar, and P. Afriwardani, "Certainty of response index-based e-diagnostics assisted by Google Forms to identify misconceptions in simple harmonic waves," *Journal of Education Research and Evaluation*, vol. 7, no. 1, pp. 168–174, Apr. 2023, doi: 10.23887/jere.v7i1.56351.

[10] K. Fenditasari, Jumadi, E. Istiyono, and Hendra, "Identification of misconceptions on heat and temperature among physics education students using four-tier diagnostic test," *Journal of Physics: Conference Series*, vol. 1470, no. 1, p. 012055, Feb. 2020, doi: 10.1088/1742-6596/1470/1/012055.

[11] D. Kaltakci Gurel, A. Eryilmaz, and L. C. McDermott, "A review and comparison of diagnostic instruments to identify students' misconceptions in science," *EURASIA Journal of Mathematics, Science and Technology Education*, vol. 11, no. 5, pp. 989–1008, Oct. 2015, doi: 10.12973/eurasia.2015.1369a.

[12] Soeharto, B. Csapó, E. Sarimanah, F. I. Dewi, and T. Sabri, "A review of students' common misconceptions in science and their diagnostic assessment tools," *Jurnal Pendidikan IPA Indonesia*, vol. 8, no. 2, pp. 247–266, Jun. 2019, doi: 10.15294/jpii.v8i2.18649.

[13] S. D. Korkmaz *et al.*, "Evaluating the gifted students' understanding related to plasma state using plasma experimental system and two-tier diagnostic test," *Journal of Education in Science Environment and Health*, vol. 4, no. 1, pp. 46–53, 2018.

[14] C. Tsui and D. Treagust, "Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument," *International Journal of Science Education*, vol. 32, no. 8, pp. 1073–1098, May 2010, doi: 10.1080/09500690902951429.

[15] S. Illene, S. Feranie, and P. Siahaan, "Create multiple-choice tests based on experimental activities to assess students' 21st century skills in the heat and heat transfer topic," *Journal of Education and Learning (EduLearn)*, vol. 17, no. 1, pp. 44–57, Feb. 2023, doi: 10.11591/edulearn.v17i1.20540.

[16] S. Utari, J. Manopo, S. Feranie, E. Prima, and B. Iswanto, "Development of standardized online test to assess the students 21st century skills," in *Proceedings of the Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019, 12 October 2019, Bandung, West Java, Indonesia*, EAI, 2020. doi: 10.4108/eai.12-10-2019.2296465.

[17] M. N. Humaidi, F. A. Triansyah, R. Sugianto, and A. R. N. Laila, "Development of a HOTS-leveled two-tier multiple choice (TTMC) test to measure student misconceptions in islamic studies," *Assyfa Journal of Islamic Studies*, vol. 1, no. 1, pp. 31–40, Jun. 2023, doi: 10.61650/ajis.v1i1.148.

[18] A. Halim *et al.*, "Development of concept maps diagnostic test for identification of students' misconceptions," in *AIP Conference Proceedings*, 2020, p. 050003. doi: 10.1063/5.0000613.

[19] T. Y. Z. Cengiz, "Development of two-tier diagnostic instrument and assess students' understanding in chemistry," *Scientific Research and Essays*, vol. 4, no. 6, pp. 626–631, 2009.

[20] R. Diani, J. Alfin, Y. M. Anggraeni, M. Mustari, and D. Fujiani, "Four-tier diagnostic test with certainty of response index on the concepts of fluid," *Journal of Physics: Conference Series*, vol. 1155, no. 1, p. 012078, Feb. 2019, doi: 10.1088/1742-6596/1155/1/012078.

[21] V. Liampa, G. N. Malandrakis, P. Papadopoulou, and D. Pnevmatikos, "Development and evaluation of a three-tier diagnostic test to assess undergraduate primary teachers' understanding of ecological footprint," *Research in Science Education*, vol. 49, no. 3, pp. 711–736, 2019, doi: 10.1007/s11165-017-9643-1.

[22] S. Sadhu, M. T. Tima, V. P. Cahyani, A. F. Laka, D. Annisa, and A. R. Fahriyah, "Analysis of acid-base misconceptions using modified certainty of response index (CRI) and diagnostic interview for different student levels cognitive," *International Journal of Science and Applied Science: Conference Series*, vol. 1, no. 2, p. 91, Aug. 2017, doi: 10.20961/ijsascs.v1i2.5126.

[23] R. A. Sadevi and I. C. Sayekti, "Identification of science misconceptions using two-tier multiple choice for elementary school grade iv students," *Profesi Pendidikan Dasar*, vol. 10, no. 1, pp. 24–37, Apr. 2023, doi: 10.23917/ppd.v10i1.4063.

[24] S. R. Putri, S. N. Hofifah, G. C. S. Girsang, and A. B. D. Nandiyanto, "How to identify misconception using certainty of response index (CRI): A study case of mathematical chemistry subject by experimental demonstration of adsorption," *Indonesian Journal of Multidiciplinary Research*, vol. 2, no. 1, pp. 143–158, Sep. 2021, doi: 10.17509/ijomr.v2i1.38738.

[25] M. D. Gall, W. R. Borg, and J. P. Gall, *Educational research: An introduction*. Longman Publishing, 1996.

[26] A. N and A. Asrizal, "Design and validity of STEM integrated physics electronic teaching materials to improve new literacy of class xi high school students," *Jurnal Pendidikan Fisika*, vol. 10, no. 3, pp. 177–192, Aug. 2022, doi: 10.26618/jpf.v10i3.7900.

[27] S. Suciati, S. Munadi, S. Sugiman, and W. D. R. Febriyanti, "Design and validation of mathematical literacy instruments for assessment for learning in Indonesia," *European Journal of Educational Research*, vol. volume-9-2, no. volume-9-issue-2-april-2020, pp. 865–875, Apr. 2020, doi: 10.12973/eu-jer.9.2.865.

[28] S. Hasan, D. Bagayoko, and E. L. Kelley, "Misconceptions and the certainty of response index (CRI)," *Physics Education*, vol. 34, no. 5, pp. 294–299, Sep. 1999, doi: 10.1088/0031-9120/34/5/304.

[29] H. Iqbal, "The development and validation of a one tier diagnostic assessment to test premedical students ' misconceptions about traumatic brain injury," M.S. thesis, Wright State University, 2016.

[30] K. Parsakia, M. Rostami, and S. M. Saadati, "Validity and reliability of digital self-efficacy scale in Iranian sample," *journal of Adolescent and Youth Psychological Studies*, vol. 4, no. 4, pp. 207–215, Jul. 2023, doi: 10.61186/jayps.4.4.207.

[31] J. Randrianarivony, J. J. Ravelomanantsoa, and N. Razanamihaja, "Evaluation of the reliability and validity of the Early childhood oral health impact scale (ECOHIS) questionnaire translated into Malagasy," *Health and Quality of Life Outcomes*, vol. 18, no. 1, p. 39, Dec. 2020, doi: 10.1186/s12955-020-01296-1.

[32] A. M. R. Tumanggor and S. Supahar, "The development of diagnostic test instrument for mathematical representation ability (PhysDTRA) in high school physics learning," *Journal for the Education of Gifted Young Scientists*, vol. 8, no. 4, pp. 1439–1455, Dec. 2020, doi: 10.17478/jegys.777425.

[33] K. S. Taber, "The use of Cronbach's Alpha when developing and reporting research instruments in science education," *Research in Science Education*, vol. 48, no. 6, pp. 1273–1296, Dec. 2018, doi: 10.1007/s11165-016-9602-2.

[34] A. Cullinane and M. Liston, *Two-tier Multiple choice question: an alternative method of formatif assessment for first year undergraduate biology students*. Limerick: National Center for Excellence In Mathematics and Education Science Teaching and Learning (NCE-MSTL), 2011.

[35] A. 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," *International Journal of Science and Applied Science: Conference Series*, vol. 4, no. 1, p. 44, Dec. 2020, doi: 10.20961/ijsascs.v4i1.49457.
[36] M. Nasir, C. Cari, W. Sunarno, and F. Rahmawati, "The development of a two-tier diagnostic test for student understanding of light," *Journal of Turkish Science Education*, vol. 20, no. 1, pp. 150–172, Mar. 2023, doi: 10.36681/tused.2023.009.
[37] P. Rintayati, H. Lukitasari, and A. Syawaludin, "Development of two-tier multiple choice test to assess indonesian elementary students' higher-order thinking skills," *International Journal of Instruction*, vol. 14, no. 1, pp. 555–566, Jan. 2021, doi: 10.29333/iji.2021.14133a.
[38] I. N. Qomariyah and M. Mistianah, "Analysis of students CRI (certainty of response index) of genetic concept based two-tier diagnostic test online using I spring suite 8 for COVID-19 learning," *Proceeding Umsurabaya*, pp. 136–145, 2020.
[39] I. Koto and S. E. Gusma, "Using certainty response index to differentiate lack of knowledge and misconception about basic electrical concepts," *Journal of Physics: Conference Series*, vol. 1731, no. 1, p. 012070, Jan. 2021, doi: 10.1088/1742-6596/1731/1/012070.
[40] Y. Hala, S. Saenab, A. Novia Arifin, and S. Satar, "Identification of senior high school student'smisconceptions in makassar city on cell concepts by using the certainty of response index (CRI) method," *Journal of Physics: Conference Series*, vol. 1317, no. 1, p. 012194, Oct. 2019, doi: 10.1088/1742-6596/1317/1/012194.
[41] F. H. Latif, Mursalin, T. J. Buhungo, and A. H. Odja, "Analysis of students' misconceptions using the certainty of response index (CRI) on the concept of work and energy in SMA Negeri 1 Gorontalo utara after online learning," in *Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)*, 2021, pp. 511–515. doi: 10.2991/assehr.k.210305.075.
[42] M. Marzuki and M. Diknasari, "Misconceptions: an analysis of certainty of response index (CRI) on photosynthesis materials for junior high school students," *Biosfer: Jurnal Tadris Biologi*, vol. 13, no. 1, pp. 49–55, Oct. 2022, doi: 10.24042/biosfer.v13i1.12480.
[43] T. Tukiyo, M. Efendi, E. M. Solissa, I. Yuniwati, and S. A. Pranajaya, "The development of a two-tier diagnostic test to detect student's misconceptions in learning process," *Mudir: Jurnal Manajemen Pendidikan*, vol. 5, no. 1, pp. 92–96, 2023, doi: org/10.55352/mudir.v5i1.33.
[44] G. Resbiantoro, R. Setiani, and Dwikoranto, "A review of misconception in physics: the diagnosis, causes, and remediation," *Journal of Turkish Science Education*, vol. 19, no. 2, pp. 403–427, 2022.
[45] Zainuddin, Mujakir, M. Ibrahim, B. Jatmiko, A. Halim, and Yusrizal, "Model of cocoaer as alternative learning to prevent the potential misconceptions of high school students," *Journal of Physics: Conference Series*, vol. 1460, no. 1, p. 012126, Feb. 2020, doi: 10.1088/1742-6596/1460/1/012126.

## BIOGRAPHIES OF AUTHORS

**Rita Arfi Astuti Ningroom** born in Karanganyar, on July 1, 1996. She is one of the students of the Postgraduate Elementary School Teacher Education Study Program, Faculty of Teacher Training and Education, Sebelas Maret University, Surakarta-Indonesia. Her research interests are science education, misconceptions in elementary school science learning, instruments in material concepts and changes, and the development of instruments to detect misconceptions in students. She participated in "The 4th International Conference of Economics and Entrepreneurship" with the theme Economics and Entrepreneurship Education Link and Match between Government, Universities and Industry" in 2020 with the article titled: "Implementation of Digital Literacy in Primary Science Learning during the COVID-19 Pandemic". She can be contacted at email: arfirita_astngr@student.uns.ac.id.

**Sri Yamtinah** born in Surakarta, on December 4, 1969. She is a lecturer in Chemistry Education at the Faculty of Teacher Training and Education, Sebelas Maret University, Surakarta-Indonesia. She completed a bachelor's degree in Chemistry Education at KIP UNS Surakarta in 1993 and completed postgraduate education in the field of educational research and evaluation in 1999 and doctoral studies in 2014 at the Department of Educational Research and Evaluation at Yogyakarta State University, Indonesia. She has compiled several scientific articles published in national and international journals. She is also the author of a book entitled learning evaluation which was published through UNS Press. She can be contacted at email: jengtina@staff.uns.ac.id.

**Riyadi** Born in Kebumen, on January 16, 1967. He is a lecturer in Elementary School Teacher Education at the Faculty of Teacher Training and Education, Sebelas Maret University, Surakarta-Indonesia. He completed a bachelor's degree in Mathematics Education at Semarang State University-Indonesia in 1993 and completed postgraduate education in 2000 and doctoral studies in 2009 in Mathematics at Gadjah Mada University, Yogyakarta-Indonesia. He has carried out several research and community services since 2007. He has also compiled several scientific articles published in national and international journals. He can be contacted at email: riyadi_pgsd_fkip@staff.uns.ac.id.