Journal of Turkish Science Education, 2025, *22*(2), 338-353.
DOI no: 10.36681/tused.2025.017

Journal of Turkish Science Education
http://www.tused.org
© ISSN: 1304-6020

Development of a four-tier diagnostic test for misconceptions in natural science of primary school pupils

Anatri Desstya[1], Ika Candra Sayekti[2], Muhammad Abduh[3], Sukartono[4]

[1]*Universitas Muhammadiyah Surakarta, Indonesia, Correspondent author, ad121@ums.ac.id, ORCID ID: 0000-0002-2643-9252*
[2]*Universitas Muhammadiyah Surakarta, Indonesia, ORCID ID: 0000-0002-6556-796X*
[3] *Universitas Muhammadiyah Surakarta, Indonesia, ORCID ID: 0000-0002-2643-9252*
[4] *Universitas Muhammadiyah Surakarta, Indonesia, ORCID ID: 0000-0002-5739-1044*

ABSTRACT

This study aimed to develop a standardised instrument for diagnosing science misconceptions in primary school children. Following a developmental research approach using the 4-D model (Define, Design, Develop, Disseminate), 100 four-tier multiple choice items were constructed. Content validity was established through expert evaluation by six science specialists. The study was conducted using stratified purposive sampling on 140 pupils from public and private primary schools with 'very good' and 'good' ratings in Surakarta, Indonesia. The instrument's construct validity, reliability, and empirical validity were assessed using Aiken's V, Pearson's Correlation Coefficient, and Cronbach's alpha, respectively. Difficulty index was also calculated. The final version comprised 61 valid and reliable items with Aiken's V ranging from 0.79 to 0.92, Pearson's Correlation ranging from 0.17 to 0.58, and a Cronbach's alpha coefficient of 0.86. Item difficulty ranged from 0.014 to 0.62, with 27 and 34 items falling into the moderate and difficult categories, respectively. This instrument demonstrates potential for effectively detecting science misconceptions in primary school learners. Based on these findings, teachers, policymakers and parents can take targeted action to address misconceptions. This may involve evaluating science textbooks, enhancing teacher training, and fostering discussions with science experts. Additionally, parental involvement in reinforcing science concepts is crucial.

RESEARCH ARTICLE

ARTICLE INFORMATION
Received:
5.10.2023
Accepted:
27.06.2024
Available Online:
30.06.2025

KEYWORDS:
Misconceptions, science, diagnostic test, aiken's V, 4-D model.

To cite this article: Desstya, A., Sayekti, I.C., Abduh, M., & Sukartono, S. (2025). Development of a four-tier diagnostic test for misconceptions in natural science of primary school pupils. *Journal of Turkish Science Education, 22*(2), 338-353. http://doi.org/10.36681/tused.2025.017

Introduction

Science constitutes a body of knowledge comprising a collection of facts, concepts, laws, principles, theories and models (Chiappeta & Koballa, 2010). Knowledge frameworks concerning learning and scientific practice are now influencing policy initiatives that have an impact on primary school science teaching (Abell & Lederman, 2007; Nugroho et al., 2021). Desstya & Sayekti (2020) stated science is one of the subjects taught in primary schools in Indonesia for several reasons. Science is universally beneficial as fundamental scientific principles drive technological advancements, facilitating easier lifestyles globally. Science dispels myths and unscientific beliefs, promoting rational understanding. Science underpins various professions by revealing logical foundations, for example, doctors rely heavily on scientific knowledge based on medical and biological research. Science fosters critical thinking and problem-solving skills, essential for intellectual development. Science instils positive characters and attitudes (refers to traits, attitudes and behavior that demonstrate good qualities) nurturing curiosity about phenomena in the surrounding world. Everyday occurrences expose primary school children to scientific phenomena, from observing light and optics, heat and temperature when boiling water. Children initially encounter science through daily experiences, which, if not aligned with accurate scientific concepts, can lead to misconceptions (Canada et al., 2017). Misconceptions denote inaccurate understandings, misapplications of concept names, or incorrect classifications of examples (Bou'-Jaoude, 1991; Ajayi, 2017; Berg, 1991; Chang, 2010). They impede knowledge development and scientific thinking (Bou-Jaoude, 1991), significantly impacting learning success (Bou-Jaoude, 1991), and hindering the learning process (Bau-Jaoude, 1991). Misconceptions are prevalent among teachers, lecturers, students (at both school and university levels), and researchers.

Numerous studies have documented science misconceptions among primary school learners (Nurjani et al., 2020; Deringol, 2019; Bar et al., 2016; Kambouri, 2010; Narjaikaew, 2013; Sahin et al., 2008; Smolleck & Hershberger, 2011; Stein et al., 2008), necessitating teachers' intervention. Addressing misconceptions promptly is crucial to remove conceptual barriers to science learning (Soeharto et al., 2019; Nurhasanah et al., 2022). Early detection of science misconceptions is accordingly essential to mitigate their negative impacts. Diagnostic tests are valuable tools for identifying misconceptions (Directorate General of Ministry of National Education; Directorate of Junior High School Development, 2007). Various diagnostic tools have been developed and employed to pinpoint science misconceptions. Soeharto et al. (2019) revealed that found that a number of concepts in physics (33 concepts), chemistry (12 concepts), and biology (15 concepts) commonly causing misconceptions. Furthermore, it found that the types of diagnostic tools used include interviews (10.74%), simple multiple-choice tests (32.23%), multiple-tier tests (33.06%), and open-ended tests (23.97%). Each diagnostic tool possesses its strengths and limitations. Multiple-tier tests, consisting mainly of multiple-choice formats, are widely used for detecting science misconceptions (Soeharto et al, 2019). Despite the efficiency of multiple-choice tests as evaluation tools (Kusumawati & Hadi, 2018), they have drawbacks. Although comprehensive and objective, multiple-choice questions suffer from limitations such as increased error variance due to guessing, lack of insight into respondents' thought processes, and constraints on answer options (Caleon & R. Subramanium, 2010; Suwarna, 2014). Multiple-choice tests including two-tier and three-tier formats, were developed to address these shortcomings and provide more nuanced assessments.

The two-tier type is a two-level multiple-choice test that presents answer choices and reasons for their selection. This instrument has been developed by several researchers (Nasir et al., 2023; Soeharto, 2021; Sadevi, R.A & Sayekti, I.C, 2023). Questions in this type of test often be unclear to distinguish between correct and incorrect answers due to conceptual difficulty or misconceptions. These questions typically necessitate thorough preparation and specialised assistance (Rintayati, Lukitasari, & Syawaludin, 2021).

The three-tier test is a multiple-choice test comprising three levels of answers: answer choices, reasons, and confidence levels in answering a question (Caleon & Subramanium, 2010; Mubarokah, Mulyani, Indriyanti, 2018). However, this test type's weakness lies in potentially low interpretation proportions (that means it is difficult to distinguish whether an incorrect answer is due to misconception, ignorance, or simply guessing, especially if the participant's level of confidence is low) for those who lack conceptual understanding and excessively high proportions for those who do understand. Consequently, distinguishing between respondents with misconceptions and those lacking conceptual understanding remains challenging (Diani et al., 2019). This challenge complicates determining appropriate corrective approaches for learners with misconceptions versus those simply lacking understanding. Hence, there is a need to develop a multilevel multiple-choice test in the form of a four-tier multiple-choice test. This four-level diagnostic test, as proposed by Maharani et al.

(2019); Salamah et al. (2020); Rohmanasari & Ermawati (2019), and Putica, B. (2022) comprises four levels: the first level containing answer choices, the second level containing the confidence level in the chosen answers, the third level containing reasons for selecting the answers in the first level, and the final level containing the confidence level in the written reasons. In this study, the investigation focuses on answers and reasons using closed options to maintain pupil engagement with the material being assessed.

Yuberti et al. (2020) developed a four-tier diagnostic test to identify misconceptions at the junior high school level. The test focuses on exploring the concepts of work and simple planes. Habiddin & Page (2019) developed a four-tier multiple-choice test for university students, concentrating on the concept of chemical reaction kinetics. Saputra et al. (2020) devised a four-tier multiple-choice test for high school students, delving into fluid mechanics concepts. These tests, developed by Yuberti et al. (2020) and Saputra et al. (2020), feature four levels, consisting of answers to the question, confidence level in answering, reasons for answering, and level of confidence for the answer/reason respectively. This diagnostic instrument offers various advantages, such as including the level of confidence in answering and the opportunity to give reasons separately (Tumanggor et al., 2020; Celikkanlı, Kızılcık, 2022), facilitating easier implementation and access (Nurhidayatullah & Prodjosantoso, 2018). Kartimi et al. (2021) developed a four-tier diagnostic test to identify primary school children's misconceptions regarding the concept of energy. Comprising six questions for fourth graders, this test contributes to alternative ways to identify and analyse misconceptions. Other advantages of this four-tier multiple-choice test include its practicality in implementation and scoring, as well as the efficiency of time and energy used (Agustin, 2020). However, the weakness of this test is that it cannot be used to determine the factors that contribute to those misconceptions. Among the existing types, the four-tier multiple choice at the first level is a multiple-choice question with four distractors and one correct answer key. The second level is the level of confidence in choosing answers. The third level is the reason for choosing an answer.

However, the four tiers have not been able to determine the sources causing misconceptions. Therefore, the development of this type of four-tier multiple-choice test requires modification to address the aspects contributing to misconceptions. Understanding the sources of error is crucial as it informs the design of effective learning interventions. This aligns with research by Inggit et al. (2021), suggesting that proper learning strategies should be tailored to address misconceptions and their underlying causes. Further research analysing factors contributing to learners' misconceptions is warranted (Sheftyawan, Prihandono, & Lesmono (2018). Previous studies have demonstrated that four-tier multiple-choice tests have been extensively developed to detect science misconceptions at middle school, high school this term means something a bit different in various systems, and tertiary levels, focusing on specific topics. While this instrument has been adapted for use at the primary school level, its usage remains limited and requires further refinement. Among the existing four tiers, none have been able to pinpoint the sources of experienced misconceptions (Salamah, 2020). Thus, modifying the format of the four-tier multiple-choice test is essential for effective use and for better identification of factors contributing to students' misconceptions. The four-tier multiple-choice test developed for primary school learners serves as a novel instrument for detecting science misconceptions. From the child's perspective, this instrument fosters critical, analytical, and reflective thinking skills. In tiers 1 and 2, they are prompted to analyse scientific situations, concepts, problems, and evaluate decisions. Meanwhile, tiers 3 and 4 encourage them to reflect on their understanding. From the teacher's standpoint, this instrument provides valuable insights into the sources of misconceptions, enabling the preparation of tailored follow-up plans for science learning.

Therefore, the development of an instrument to detect science misconceptions in the form of a four-tier multiple-choice test is essential due to its numerous benefits. This test can identify misconceptions accurately, allowing pupils to determine answers to explored concepts, provide reasons for their choices, demonstrate confidence in their responses, and identify the sources of their answers. Hence, this study aims to develop such an instrument that meets the requirements of validity, reliability, and difficulty index. Consequently, this research addresses the following questions: a) How are the characteristics of the Four-tier Diagnostic Test for Misconceptions; (b) How is the validity and reliability based on experts assessment; c) How are the validity, reliability and difficulty level of the instrument in empirical tests?

Methods

Participants

The participants in this study were 140 fifth-grade primary school pupils aged 11 to 12 years, comprising 76 boys and 64 girls. The sample was obtained through stratified purposive sampling, involving public and private schools with 'very good' and 'good' quality school accreditation (Cohen, Manion, & Morrison, 2018), in five sub-districts in Surakarta, Central Java, Indonesia, to ensure the representation of various sample characteristics.

Development Framework

This research follows a Research and Development (R&D) approach. The development model used was the 4-D model developed by Thiagarajin, Semmel & Senmel (1974), which consists of defining, designing, developing and disseminating stages. The 4-D model was chosen for its simplicity in the product development process (Irawan et al., 2017), and its detailed steps allow for systematic use. The defining stage involved the analysis of pupil characteristics, misconceptions and available diagnostic instruments. During the design stage, the instrument was developed by devising constructs for each question item, determining core competencies and basic competencies, and developing indicators of competency achievement. The developing stage involved developing the instrument for expert review, with all validated and revised items then tested on pupils. In the dissemination stage, the instrument for identifying science misconceptions was disseminated to primary school teachers.

Instrument

Question Item Constructs

The questions were based on the standard content of the fifth-grade science curriculum, covering the following topics: locomotory organs of animals and humans and their functions, respiratory organs of animals and humans and their functions, digestive organs of animals and humans and their functions, circulatory organs of animals and humans and their functions, the relationship between ecosystem components and food webs in the surrounding environment, heat transfer, effect of heat on changes in temperature and shape of objects, the water cycle and its impact on events on Earth and the survival of living things, and matter in daily life based on its constituent components (single and mixed substances). Ten to eleven questions were developed for each topic.

Item Format

Initially, this misconception detection instrument consisted of 100 four-tier multiple-choice questions. The first tier presented answer choices for the concept asked, the second tier provided reasons for choosing the answer, the third tier indicated the level of confidence in answering, and the fourth tier identified the source of the answer. Before constructing question items, it was necessary to create a content outline based on basic competency and competency achievement indicators that had been developed.

Scoring Procedure

The maximum score for correct answers in each tier is 1.

Data Analysis Technique

To ensure the validity and reliability of the instrument for detecting misconceptions in science, the Aiken's V formula (Aiken, 1985) was employed for validity analysis. The formula is as follows:

$$V = \frac{S}{[n(C-1)]} \qquad (1)$$

Where :
S : R - lo
V : Validity Index
R : The Score Given by The Experts
lo : Lowest Validity Score
c : Highest Validity Score
n : Total Number of Experts

The validity criteria for this instrument were determined based on the number of validators involved in the study. Six validators participated, including one professor specialising in physics education, two individuals with doctoral degrees in research and evaluation in chemistry and biology education, and three individuals with Master's degrees in physics and biology education. The validation sheet included multiple categories (five in this study), and the significance level was set at $p < 0.05$. According to these criteria, the V-value was established at 0.79. Items with a minimum V of 0.79 were empirically tested to assess validity between items. Validity calculations were performed using SPSS 23 software, determining bivariate correlations between item scores and the total construct score. An item was considered valid if the bivariate correlation score exceeded 0.5 (Perinetti, 2019).

Reliability refers to the consistency and trustworthiness of a test instrument, which is deemed reliable if it produces consistent results under identical conditions. Reliability was assessed by administering the instrument to 60 respondents. Cronbach's alpha method was employed to test instrument reliability, conducted using SPSS 23 software. The results indicated high reliability, with a coefficient exceeding 0.7 (Gelişli et al., 2017).

The difficulty index was established with the following ranges: 0.80 to 1.00 (easy), 0.30 to 0.79 (moderate), and 0.00 to 0.29 (difficult). The formula used to calculate the difficulty index is as follows (Jandaghi, 2011; Jandaghi & Shaterian, 2008).

$$p = \frac{B}{JS} \qquad (2)$$

Where:
p: The difficulty Index
B: The number of student who answered correctly
JS: The number of student taking the test

To categorise respondents' answers into the category of misconceptions or not, the grouping criteria based on the three-tier multiple-choice test outlined by Gurel, Eryilmaz, and McDermott (2015) were utilised. These criteria, presented in Table 1, were selected because the fourth tier in the multiple-choice test developed in this study pertains to identifying the source of answers chosen in the previous tiers or to pinpoint the source of misconceptions.

Table 1

Criteria for grouping students'answer choices

First Tier	Second Tier	Third Tier	Category
Correct	Correct	Confident	Understand the concept
Correct	Incorrect	Confident	Misconception *(false positive)*
Incorrect	Correct	Confident	Misconception *(false negative)*
Incorrect	Incorrect	Confident	Misconception
Correct	Correct	Unconfident	Consecutive guess, Lack of confidence
Correct	Incorrect	Unconfident	Lack of concept understanding
Incorrect	Correct	Unconfident	Lack of concept understanding
Incorrect	Incorrect	Unconfident	Lack of concept understanding

Note. Source from Gurel et. Al (2017) with author's modification.

Findings

Table 2 shows distribution of valid and invalid questions

Table 2

Distribution of valid and invalid questions

Evaluation of question quality	Question Number	Numbers of Questions
Valid	2,3,4,5,6,7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 38, 39, 41, 42, 45, 47, 48, 50,51, 53, 54, 55, 57, 58, 59, 61, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 75, 77, 78, 79, 80, 81, 82, 84, 85, 87, 89, 90, 92, 93, 96, 98, 99.	73
Invalid	1, 16, 20, 21, 22, 27, 33, 40, 43, 44, 46, 49, 52,56, 59, 60,62, 74, 76, 83, 86, 88, 91,94, 95, 97, 100	27
Total		100

The Aiken's V coefficient for valid questions ranged from 0.792 to 0.917, with an average score of 0.793. Subsequently, after identifying 73 valid questions based on experts' judgment, these questions were administered to a sample of primary school children. The test results yielded scores ranging from 2 to 4. These scores were utilised to calculate the bivariate correlation between each question's score and the overall total score using the SPSS program. The correlation values are presented in Table 3.

Table 3

Pearson's correlation value of each item

Items	Pearson's Correlation	Result	Item Number	Pearson's Corelation	Result	Item Number	Pearson's Corelation	Result
1	0.206*	Valid	25	0.446**	Valid	49	0.092	Invalid
2	0.321**	Valid	26	0.421**	Valid	50	0.500**	Valid
3	0.331**	Valid	27	0.182*	Valid	51	0.043	Invalid
4	0.437**	Valid	28	0.403**	Valid	52	0.513**	Valid

Items	Pearson's Correlation	Result	Item Number	Pearson's Corelation	Result	Item Number	Pearson's Corelation	Result
5	0.308**	Valid	29	0.462**	Valid	53	0.199*	Valid
6	0.142	Invalid	30	0.576**	Valid	54	-0.288**	Valid
7	0.262**	Valid	31	0.213*	Valid	55	0.392**	Valid
8	0.431**	Valid	32	0.457**	Valid	56	0.023	Invalid
9	0.068	Invalid	33	0.331**	Valid	57	0.526**	Valid
10	0.194*	Valid	34	0.422**	Valid	58	0.356**	Valid
11	0.182*	Valid	35	0.535**	Valid	59	0.440**	Valid
12	-0.021	Invalid	36	0.274**	Valid	60	0.284**	Valid
13	-0.072	Invalid	37	0.230**	Valid	61	0.054	Invalid
14	0.189*	Valid	38	0.310**	Valid	62	0.514**	Valid
15	0.221**	Valid	39	-0.038	Invalid	63	0.352**	Valid
16	0.230**	Valid	40	0.276**	Valid	64	0.295**	Valid
17	0.168*	Valid	41	0.254**	Valid	65	0.215**	Valid
18	0.394**	Valid	42	0.368**	Valid	66	0.407**	Valid
19	0.338**	Valid	43	0.296**	Valid	67	0.480**	Valid
20	0.142	Invalid	44	0.100	Invalid	68	0.240**	Valid
21	0.363**	Valid	45	0.112	Invalid	69	0.311**	Valid
22	0.410**	Valid	46	0.461**	Valid	70	0.255**	Valid
23	0.441**	Valid	47	0.274**	Valid	71	0.457**	Valid
24	0.498**	Valid	48	0.500**	Valid	72	0.325**	Valid

Based on Table 3, it is evident that out of the 73 questions, 61 were deemed valid, indicating their suitability for measuring the intended competency (Nunnally, 1978). However, 12 questions were deemed invalid. The criterion for validity determination was set at a minimum score of 0.164, calculated using the two-tailed r table with a significance level of 5% and 140 samples. The valid items ranged in scores from 0.168 to 0.576.

Reliability

Cronbach's alpha was found to be 0.863, indicating high reliability.

Difficulty Index of Questions

From the study results, among the 61 valid and reliable questions, 27 fell into the "moderate" category with scores between 0.30 and 0.69, while 34 questions were categorized as "difficult" with scores ranging from 0.00 to 0.29. None of the questions were classified as "easy." Consequently, eight questions required correction and modification to enhance their quality and acceptability. The responses to select questions among the 61 valid and reliable items were scrutinised to identify any misconceptions, and the findings are elaborated upon in the subsequent section.

Disseminating

During the dissemination phase, the tool for pinpointing misconceptions in science was distributed to 38 primary school teachers through instructional workshops and mentoring sessions aimed at addressing science misconceptions in their pupils. These teachers were selected from various public and private schools across five sub-districts (Banjarsari, Serengan, Pasarkliwon, Jebres, Laweyan) including Gading Public Primary School, Islamic Primary School Sunan Kalijaga, Mijipinilihan Public Primary School, Muhammadiyah 10 Primary School, Islamic Bakti 1 Primary

School, Baturono Public Primary School, Kleco 2 Public Primary School, Mojosongo 2 Public Primary School, Al Firdaus, Ar Risalah, Kemasan 1 Public Primary School, and Islamic Sunniyah Primary School. The selection process employed purposive random sampling.

Schematically, the entire process of developing this item is presented in Table 4.

Table 4

Stage of the item development process

Stage	Process	Results
1. Initial Developed 100 item	Expert validation, and validity analysis using Aiken V	73 valid items 27 non valid items
2. Continued 73 valid items	Tested on 140 students, validity analysis (used *Pearson's Correlation)*, reliabilitas (by coefficient of Cronbach's alpha)	61 valid items 12 non valid items
3. Final Obtained 61 valid items	difficulty index.	61 valid items (27 questions in the "moderate" and 34 questions in the "difficult" category)

Discussion

Content validity assesses the extent to which elements in a measurement instrument are relevant and representative of the construct being measured (Bahri, 2019; Ishartono et al. 2024). It is a pivotal step in instrument development, as it precedes the testing of construct validity. An instrument must pass the content validity test before undergoing other validity assessments (Ihsan, 2015). Following the validation and reliability assessment of the instrument, it was utilized to detect student misconceptions. For instance, the analysis of responses to question number 1 is presented below.

Item 1 as presented in Table 5 is an example of a valid item in this study.

Table 5

Example of item 1

Irwan works as a porter at a market. Every day he lifts many heavy sacks of rice and vegetables. One day, Irwan's arm muscles got tight and swollen, so Irwan's friend had to help him lift a sack of rice. The muscle disorder that Irwan suffered from is called… a. stiff b. atrophy c. hypertrophy d. tetanus
Reason a. The muscle disorder occurred because Irwan did continuous and excessive activities b. The muscle disorder occurred because Irwan had poliomyelitis c. The muscle disorder occurred because Irwan made a sudden stomping motion d. The muscle disorder occurred because Irwan never had exercise
Your level of confidence in answering a. confident b. unconfident
The source of your answer

a. I have ever helped/cared for a friend with the same condition as Irwan b. I have experienced the same condition as Irwan c. I have seen other people experience the same thing as Irwan d. I have seen a video of a person having the same problem as Irwan

Based on Table 5, it is evident that Sample Item 1 aims to diagnose whether there are misconceptions among students regarding the understanding of locomotory organs of animals and humans, their functions, and the maintenance of human locomotory organs' health. In this topic, the developed competency involves identifying symptoms and bone abnormalities in humans. The question presents a discourse and assesses students' understanding of muscle disorders. The student answered the question in the following order: B. Atrophy, A. The muscle disorder occurred because Irwan did continuous and excessive activities, A. Confident, and B. I have experienced the same condition as Irwan. According to the reference category of students' understanding presented in Table 5, this student is included in the category of misconception (false negative) because the answer in the first tier is wrong, the answer in the second tier is correct, and he/she answered the question confidently. Further analysis revealed that the student's misconception stemmed from personal experience. This finding aligns with research by Lin et al (2016), which suggests that personal experience can contribute to students constructing their own knowledge. However, some of the students' daily experiences can lead to misconceptions, as observed in a study conducted by Mutsvangwa (2020). Daily experiences form students' initial knowledge, which may not always be accurate due to the variability and inaccuracies in the sources of students' information and experiences.

Table 6

Example of item 36

To make a sharp and beautiful kris, a kris blacksmith needs to design, combine the elements of beauty, and carve it. In order to make a good and artistic kris, the blacksmith burns the alloy into the flaming fire to make it heated, so it can be carved. The alloy becomes heated because of... a. transmission b. radiation c. convection d. conduction
Reason a. heat travels through the metal b. heat travels through the air c. particles in the metal move from one end to the other d. heated over the fire
Your level of confidence in answering… a. confident b. unconfident
The source of your answer… a. teacher's explanation b. a story book or textbook c. I saw a metal being burned / a kris making myself d. I listened to my mother's/father's/sibling's story

As indicated in Table 6, Item 36 serves to diagnose whether pupils harbour misconceptions when appraising the effect of heat on changes in temperature and shape of objects in daily life. In the first tier, the majority of pupils selected option 'D) conduction'. The question scenario presents a situation where a kris blacksmith heats an alloy in a flaming fire to facilitate carving, with the correct answer being ‘D. conduction’. In the second tier, the correct answer is identified as A (heat travels through metal). The responses from the first to the third tier are correct, incorrect, and confident, respectively. Consequently, it falls into the category of misconceptions (false positive). In the fourth tier, it is evident that the source of the misconception is attributed to the teacher’s explanation.

Table 7

Example of item 16

Every human should try to keep their digestive organs healthy by consuming healthy and nutritious food so that the digestion process can run well. In the human digestive system, absorption of nutrients in food occurs in the… a. mouth b. oesophagus c. stomach d. small intestine
Reason a. The process of breaking down of food and absorption of nutrients in food occurs mechanically and chemically in the mouth b. The process of breaking down of food and absorption of nutrients in food occurs mechanically and chemically in the oesophagus c. The process of breaking down of food and absorption of nutrients in food occurs mechanically and chemically in the stomach d. The process of breaking down of food and absorption of nutrients in food occurs mechanically and chemically in the small intestine
Your level of confidence in answering… a. Confident b. Unconfident
The source of your answer… a. teacher’s explanation b. a story book or textbook c. I saw a video d. I listened to my mother’s/father’s/sibling’s story

As delineated in Table 7, item 16 is designed to identify misconceptions pertaining to the digestive organs of animals and humans, their functions, and methods to maintain their health. In the first tier, the correct response to the question, “Every human should strive to maintain the health of their digestive organs by consuming nutritious food, thereby facilitating efficient digestion. In the human digestive system, the absorption of nutrients from food occurs in the…” is ‘D (small intestine)’. For the rationale (second tier), the correct response is ‘D (the process of food breakdown and nutrient absorption occurs both mechanically and chemically in the small intestine)’. It can be observed that the pattern of the response is incorrect, incorrect, and confident, indicating a misconception. The source of the misconception is identified in the fourth tier (C. I saw a video).

This instrument was utilised to assess the level of understanding of 140 elementary school pupils, who were subsequently categorised into aspects: understanding concepts, misconceptions, guessing, and lack of understanding concepts. The results revealed that 11 % (n=16) demonstrated a clear understanding of the concept, while 34% (n=47) exhibited misconceptions. A further 12% (n=7) appeared to be guessing, and the remaining 43% (n=60) showed a lack of understanding of the concept. In addition, the study identified specific misconceptions among the pupils, categorised by the topic of the material. These findings are succinctly presented in Table 8. This research thereby contributes to a more nuanced understanding of the factors influencing the comprehension of digestive health among students.

Table 8

Summary of students' misconceptions

No	Topics	Percentage (%)
1	Locomotory organs of animals and humans and their functions	23.51
2	Respiratory organs of animals and humans and their functions	30.44
3	Digestive organs of animals and humans and their functions	43.24
4	Circulatory organs of animals and humans and their functions	55.55
5	Relationships between ecosystem components and food webs	32.33
6	Heat transfer concept	24.33
7	The effect of heat on changes in temperature and shape of objects	24.33
8	Water cycle and its impact on events on earth and the survival of living things	15.34
9	Matter in daily life based on its constituent components (single and mixed substances)	50.44

Table 8 shows that the topic of animal and human circulatory organs and their functions has the highest percentage of misconceptions. This highlights that the concept is complex and requires greater attention in the learning process. For instance, common misconceptions include misunderstandings about the role of the heart, blood vessels, and the differences between blood circulation in humans and animals, as also reported by Setiabudi, Mulyadi, Puspita (2019). The instrument developed in this study has a reliability score of 0.863, indicating a very high level of consistency, making the measurement results trustworthy. High reliability in such instruments is crucial for identifying misconceptions, as also emphasized by Sadler, et al. (2019). They developed diagnostic tools for science concepts based on mental models and reported a reliability threshold above 0.80 as the ideal standard for internal validity.

The results of this study are in line with Caleon & Subramaniam (2017), which developed diagnostic instruments on human physiology material, such as the blood circulatory system, which is often a source of misconceptions due to understanding the contents of textbooks or less interactive learning media. This research recommends the use of visual media and a simulation-based approach to minimize student misconceptions. The results of this research have significant implications for the development of teaching materials for teachers. Teachers need special training to detect student misconceptions early on, especially at the primary school level. This is as suggested by Furtak et al. (2020), who emphasized the importance of misconception-based scaffolding in improving the quality of science learning. Treagust (2018) also believes that in developing assessment for science subjects, they should not only be relied on in terms of content but also explore students' misconceptions comprehensively. Treagust emphasized that the use of multiple-choice diagnostic tests designed specifically for abstract concepts, such as the circulatory system, can help teachers understand thinking patterns and design more effective teaching strategies.

The four-tier diagnostic test instrument developed in this study is believed to be capable of identifying misconceptions and their underlying causes among primary school students. This instrument has four tier explored concept, provide reasons for their choices, demonstrate confidence in their responses, and identify the sources of their answers. This statement aligns with Gurel et al. (2015), highlighted that assessment tools are particularly effective in thoroughly identifying misconceptions. Similarly, Kaltakci-Gurel et al. (2016) developed a four-tier assessment focused on electricity in physics, discovering that it successfully revealed misunderstandings stemming from the misinterpretation of textbook material and poor teaching practices. This underscores the necessity for educational resources that can enhance students' conceptual comprehension. Treagust et al. (2017) further validate the usefulness of diagnostic assessments in the early detection of student misconceptions. According to Treagust, utilizing four-tier assessments encourages students to engage in analytical thinking, particularly regarding intricate scientific subjects such as human body systems or changes in matter. The advancement of this assessment tool carries important implications for science education at the primary level, particularly in improving instructional effectiveness through the early identification of misconceptions. This supports Cheung et al. (2017), who proposed that four-level test evaluations can help bridge the gaps in students' misunderstandings.

For primary school teachers, instruments produced in this research is considered comprehensive. They must be prepared to identify the types of understanding or misconceptions of their pupils. If students exhibit misconceptions, the teacher can immediately identify the source and address it. The approach to this depends on each factor. If the source of the misconception is the teacher’s explanation, then the teacher must design learning that applies models, approaches, methods, and strategies that align with the characteristics of the material. Moreover, teachers must master the material thoroughly, not only by exploring the material independently but also through various professional activities, such as scientific activities or forums, as a suggested by Mitkovska (2010). Essentially, teachers are required to continually broaden their horizons to ensure the effective delivery of science learning in the classroom. When considering factors such as a storybook or textbook, watching a video, or listening to a parent’s or sibling’s story, parental involvement at home is crucial. Parents must accompany children in reading books or watching videos, ensuring the story’s plot is coherent, or providing insight into the pictures in the accompanying book. The same applies to storytelling to children. Parents should narrate phenomena in a comprehensive and uninterrupted manner. This demonstrates that parents’ basic knowledge significantly influences the misconceptions experienced by students, as conveyed by Lestari (2019). If parents possess a good basic knowledge of science material, they will be competent and capable of providing explanations and assistance to their children. They serve as their children’s teachers at home. This underscores the need for parental involvement in understanding science concepts.

Conclusion and Implications

The findings of this study indicate that the Four-tier Diagnostic Test for Misconceptions comprised 61 valid and reliable items with Aiken's V ranging from 0.79 to 0.92, Pearson's Correlation ranging from 0.17 to 0.58, and a Cronbach's alpha coefficient of 0.86. Item difficulty ranged from 0.014 to 0.62, with 27 and 34 items falling into the moderate and difficult categories, respectively. This instrument demonstrates potential for effectively detecting science misconceptions in primary school learners. The process involved the analysis of scientific situations, concepts, and problems, the evaluation of decisions, and the identification of the sources causing misconceptions. In conclusion, teachers and policymakers can undertake several measures to address the sources contributing to misconceptions. These may include evaluating the textbooks used in teaching, enhancing the teaching skills of teachers, or facilitating more frequent discussions with individuals who possess a deeper understanding of the topics. This study underscores the necessity of parental involvement in understanding science concepts. It recommends further research to investigate the factors causing misconceptions from various aspects.

References

Abell, S. K., & Lederman, N. G. (2007). Handbook of research on science education. in measurement *(Issue 801). Routledge & Francis Group*. https://doi.org/10.4324/9780203097267.ch3

Agustin, U. (2020). Development of a four-tier multiple-choice diagnostic test instrument for identifying student misconceptions in the topic of chemical equilibrium. *Semarang State University*. https://doi.org/10.15294/chemined.v11i1.40935

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/00131644854 51012

Ajayi, V. O. (2017). Misconceptions. in trends, theory and practice in science education research *(Issue October*). https://www.researchgate.net/publication/320172303_Misconceptions#fullTextFileContent

Bahri, M. F., & S. (2019). Content validity and reliability analysis of integrated islamic-science test instrument to measure the student's critical thinking ability. *Islam Realitas: Journal of Islamic & Social Studies, 5*(1), 45–51. https://doi.org/10.30983/islam_realitas.v5i1.894

Bar, V., Brosh, Y., & Sneider, C. (2016). Weight, mass, and gravity : threshold concepts in learning ccience. *Science Educator, 25*(1), 22–34. https://www.researchgate.net/publication/303523909_Weight_Mass_and_Gravity_Threshold_Concepts_in_Learning_Science/citations#fullTextFileContent

Bou-Jaoude, S. (1991). A study of the nature of students' understandings about the concept of burning. *Journal of Research in Science Teaching, 28,* 689–704. http://dx.doi.org/10.1002/tea.3660280806

Berg, E. van D. (1991). Physics misconceptions and remediation. *Satyawacana University.* https://onesearch.id/Record/IOS4813.INLIS000000000006720

Caleon, I.S, & Subramanium, 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/10.1080/09500690902890130

Caleon, I. S., & Subramaniam, R. (2017). Do students know what they know and what they don't know? using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Res Sci Educ*, *40*(3), 313-337. https://link.springer.com/article/10.1007/s11165-009-9122-4#citeas

Canada et all. (2017). Change in elementary school students' misconceptions on material systems after a theoretical-practical instruction. *International Electronic Journal of Elementary Education, 9*(3), 499–510. https://www.iejee.com/index.php/IEJEE/article/view/173/169

Celikkanlı, N.O, Kızılcık, H. S (2022). A review of studies about four-tier diagnostic tests in physics education. *Journal of Turkish Science Education*, *19*(4), 1291-1311. https://doi.org/10.36681/tused.2022.175

Chang, C. Y., Yeh, T. K., & Barufaldi, J. P. (2010). The positive and negative effects of science concept tests on student conceptual understandig. *International Journal of Science Education*, *32*(2), 265–282. https://doi.org/10.1080/09500690802650055

Cheung, D., Slavin, R. E., & Lake, C. (2017). Effective programs for elementary science: a best-evidence synthesis. *Journal of Research in Science Teaching, 54*(3), 295-304. https://doi.org/10.1002/tea.21360

Chiappeta, E. L., & Thomas R Koballa, J. (2010). Science instruction in the middle and second schools: developing fundamental knowledge and skill (K. V. Canton (ed.); 7th ed.). *Allyn and Bacon.* https://www.worldcat.org/title/science-instruction-in-the-middle-and-secondary-schools-developing-fundamental-knowledge-and-skills/oclc/862090622

Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education (8th ed.). *Routledge.* https://doi.org/10.4324/9781315456539

Deringol. (2019). Misconceptions of primary school students about the subject of fractions. *International Journal Ofs Evaluation and Research in Education (IJERE)*, *8*(3), 29–38. http://doi.org/10.11591/ijere.v8i1.16290

Desstya, A; Sayekti, I. C. (2020). Elementary science learning. *Muhammadiyah University Press*. Surakarta.

Diani, R., Alfin, J., Anggraeni, Y. M., Mustari, M., & Fujiani, D. (2019). Four-tier diagnostic test with certainty of response index on the concepts of fluid. *Journal of Physics: Conference Series, 1155*(1). https://doi.org/10.1088/1742-6596/1155/1/012078

Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2020). Scaffolding scientific understanding: the role of addressing misconceptions. *Educational Researche*r, *49*(6), 398-407. https://doi.org/10.1080/09500693.2011.604684

Habiddin, & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, *19*(3), 720–736. https://doi.org/10.22146/ijc.39218

Ihsan, H. (2015). Content Validity of research measuring tools: concepts and assessment guidelines. pedagogia: *Jurnal Ilmu Pendidikan, 13*(3), 173. https://doi.org/10.17509/pedagogia.v13i3.6004

Irawan et al. (2017). Instructional Materials development through 4d model. *SHS Web of Conferences 42*, 00086 (2018). https://doi.org/10.1051/shsconf/20184200086

Ishartono, N., Razak, R.A., Halili, S.H., Jufriansyah, A., (2024). Factors shaping Indonesian preservice math teachers' digital media adoption in online mathematics teaching practice: an instrument development and validation study. *Journal on Mathematics Education. 15*(4). 1219-1250. https://doi.org/10.22342/jme.v15i4.pp1219-1250

Jandaghi, G. (2011). Assessment of validity, reliability and difficulty indices for teacher-built physics exam questions in first year high school. *Arts and Social Sciences Journal, 16*(1), 1–4. https://eric.ed.gov/?id=EJ911907

Jandaghi, G., & Shaterian, F. (2008). Validity, reliability and difficulty indices for instructor-built exam questions. *Journal of Applied Quantitative*, *3*(2), 151–155. https://files.eric.ed.gov/fulltext/EJ803060.pdf

Mitkovska, S.J. (2010). The need of continuous professional teacher 283 development. *Procedia Social and Behavioral Sciences*, *2*(2), 2921–2926. https://doi.org/https://doi.org/10.1016/j.sbspro.2010.03.441

Kaltakci-Gurel, D., Eryilmaz, A., & McDermott, L. C. (2016). Development and application of a four-tier test to assess pre-service physics teachers' misconceptions about geometrical optics. *Research in Science Education, 46*(2), 239-261. https://doi.org/10.1007/s11165-014-9461-8

Kambouri, M. (2010). Teachers and children's misconceptions in science'. *British Educational Research Association Annual Conference*, *September*, 1–22. https://doi.org/doi:10.13140/2.1.4012.3845

Kartimi et all. (2021). A Four-tier Diagnostic instrument: an analysis of elementary student misconceptions in science topic. *Journal of Research in Science Education*. 7 (Special Issue). https://doi.org/10.29303/jppipa.v7iSpecialIssue.1022

Kusumawati & Hadi. (2018). An analysis of multiple choice questions (mcqs): item and test statistics from mathematics assessments in senior high school. *Research and Evaluation in Education, 4*(1). https://doi.org/https://doi.org/10.21831/reid.v4i1.20202

Lestari, M. D, A. (2019). Science Misconceptions in biology in fifth grade elementary school teachers. Thesis. Faculty of Teacher Training and Education, *Sanata Dharma University,* Yogyakarta

Lin, Jing-wen; Yen, Miao-Hsuan; Lia, Jia-Chi; Chiu, Mei-Hung; Guo, C.-J. (2016). Examining the factors that influence students ' science learning processes and their learning outcomes : 30 years of conceptual change research. *Eurasia Journal of Mathematics, Science & Technology Education, 12*(9), 2617–2646. https://doi.org/10.12973/eurasia.2016.000600a

Maharani et all. (2019). Diagnostic Test with four-tier in physics learning: case of misconception in newton's law material. *Journal of Physics: Conference Series, 155*(012022), 1–8. https://doi.org/10.1088/1742-6596/1155/1/012022

Ministry of National Education Directorate General; Directorate of Middle School Development. (2007). Diagnostic_test. ministry of education. *Jakarta.*

Mubarokah, F.D, Mulyani, S, Indriyanti, N.M. (2018). Identifying students' misconceptions of acid-base concepts using a three-tier diagnostic test: a case of Indonesia and Thailand. *Journal of Turkish Science Education. 15 (special issue),* 51-58. https://doi.org/10.36681/

Mutsvangwa, A. (2020). A study of student teachers misconceptions on uniform circular motion. *Journal of Physics*, *1512*(1513), 1–6. https://doi.org/10.1088/1742-6596/1512/1/012029

Narjaikaew, P. (2013). Alternative Conceptions of primary school teachers of science about force and motion. *Procedia - Social and Behavioral Sciences, 88*, 250–257. https://doi.org/10.1016/j.sbspro.2013.08.503

Nasir, et all (2023). The development of a two-tier diagnostic test for student understanding of light. *Journal of Turkish Science Education*. *20*(1), 150-172. https://doi.org/10.36681/tused.2023.009

Nugroho, C.Y.R., Hariyatmi, Supriyanto. (2021). Improving science learning outcomes using experimental methods for grade IV students in the even semester of SDN 01 Tawangmangu in 2020/2021. *Journal of Education Researsch. 3*(4). 37-44. https://doi.org/10.36654/educatif.v3i4.80

Nunnally, J. C. (1978). Psychometric Theory. *Mc Graw Hill Book Company.* https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=1867797

Nurhasanah et al., (2023). Is sipoert an innovative learning? implementation and identification of students' conception on simple harmonic motion. I*ndonesian Journal on Learning and Advanced Education. 5*(1), 78-91. https://doi.org/10.23917/ijolae.v5i1.19567

Nurhidayatullah, N., & Prodjosantoso, A. K. (2018). Misconceptions about buffer solutions. *Jurnal Inovasi Pendidikan IPA*, 4(1), 41–51. https://doi.org/10.21831/jipi.v4i1.10029

Nurjani et al. (2020). Identifying the misconception of sound concepts among grade v students at SDN 192 Pekanbaru. *Journal of Physics: Conference Series, 1–5*. https://doi.org/doi:10.1088/1742-6596/1655/1/012074

Perinetti, G. (2019). Statips part vi: bivariate correlation. *South European Journal of Orthodontics and Dentofacial Research, 6*(1), 1–5. https://doi.org/10.5937/sejodr6-21664

Putica, K. B (2022). Development and validation of a four-tier test for the assessment of secondary school students' conceptual understanding of amino acids, proteins, and enzymes. *Research in Science Education*. https://doi.org/10.1007/s11165-022-10075-5

Rintayati, P., Lukitasari, H., & Syawaludin, A. (2021). Development of two-tier multiple choice test to assess indonesian elementary students higher -order thinking skills. *International Journal of Instruction, 14*(1), 555–566. https://doi.org/https://doi.org/10.29333/iji.2021.14133a

Rohmanasari1 & Ermawati, F.U (2019) Using four-tier multiple choice diagnostic test to identify misconception profile of 12th grade students in optical instrument concepts. *IOP Publishing Ltd Journal of Physics: Conference Series. 1491* 012011. https://doi.org/10.1088/1742-6596/1491/1/012011

Sadevi, R.A & Sayekti, I.C. (2023). Identification of science misconceptions using two-tier multiple choice for elementary school grade iv students. *Profesi Pendidikan Dasar. 10*(1), 24-37. https://doi.org/10.23917/ppd.v10i1.4063

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2019). The reliability and validity of diagnostic tests in science education. *Journal of Research in Science Teaching, 56*(3), 362-383.

Sahin, C., Ipek, H., & Ayas, A. (2008). Students' understanding of light concepts primary school: A cross-age study. *Asia-Pacific Forum on Science Learning and Teaching, 9*(1), 1–19. http://www.ied.edu.hk/apfslt/download/v9_issue1_files/sahin.pdf

Salamah, U, et.all (2020). Development of a four-tier multiple choice diagnostic test to detect understanding of cognitive concepts on catabolism material in xii grade at senior high school students. *Bioeduca: Jurnal Pendidikan Biologi* *2*(1). 30 – 41. https://doi.org/10.21580/bioeduca.v2i1.5997

Saputra, O., Setiawan, A., Rusdiana, D., & Muslim. (2020). Analysis of students' misconception using four tier diagnostic test on fluid topics. *International Journal of Advanced Science and Technology*, *29(*1), 1256–1266. Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/3617

Setiabudi, A; Mulyadi; Puspita, H. (2019). An analysis of validity and reliability of a teacher-made test (case study at xi grade of sman 6 Bengkulu). *Journal of English Education and Teaching (JEET), 3*(4), 522–532. https://doi.org/10.33369/jeet.3.4.522-532

Sheftyawan, W. B., Prihandono, T., & Lesmono, A. D. (2018). Identifying student misconceptions using a four-tier diagnostic test on optical geometry material. *Jurnal Pembelajaran Fisika, 7*(2), 147–153. https://doi.org/10.19184/jpf.v7i2.7921

Smolleck, L., & Hershberger, V. (2011). Playing with science: an investigation of young children's science conceptions and misconceptions. *Current Issues in Education, 14*(1), 1–32. https://www.semanticscholar.org/paper/Playing-With-Science%3A-An-Investigation-of-Young-and-Smolleck/6f38f2971c5dc7eb6b69f40a83b3483c0fc28429

Soeharto, Csapo, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students common misconception in science and their diagnostic assesment tools. *Jurnal Pendidikan IPA Indonesia, 8(*2), 247–266. https://doi.org/10.15294/jpii.v8i2.18649

Soeharto. (2021). Development of a diagnostic assessment test to evaluate science misconceptions in terms of school grades: a rasch measurement approach. *Journal of Turkish Science Education*. *18*(3), 351-370. https://doi.org/10.36681/tused.2021.78

Stein, M., Larrabee;, T. G., & Charles R. Barman. (2008). A study of common beliefs and misconceptions in physical science. *Journal of Elementary Science Education, 20*(2), 1–11. https://files.eric.ed.gov/fulltext/EJ798575.pdf

Suwarna, I. (2014). Analysis of class x students' misconceptions on physics subject material through CRI (certainty of respone index) termodifikasi. *Jurnal Lemit.* http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf

Thiagarajin, S., Semmel, D., & Senmel, M. (1974). Instructional development for training teachers of exceptional children: a sourcebook. (Issue Indiana: Indiana University). https://files.eric.ed.gov/fulltext/ED090725.pdf

Treagust, D. F. (2018). The importance of developing diagnostic instruments for science misconceptions. *International Journal of Science Education, 40*(12), 1483-1503

Treagust, D. F., Duit, R., & Fischer, H. E. (2017). Multiple representations in physics education: models and modeling. *Springer International Publishing.* http://ndl.ethernet.edu.et/bitstream/123456789/23230/1/David%20F.%20Treagust_2017.pdf

Tumanggor, A. M. R., Supahar, Kuswanto, H., & Ringo, E. S. (2020). Using four-tier diagnostic test instruments to detect physics teacher candidates' misconceptions: case of mechanical wave concepts. *Journal of Physics: Conference Series, 1440*(1), 1–8. https://doi.org/10.1088/1742-6596/1440/1/012059

Wulandari, D., Susanti, S., & Purwanti, E. (2021). Development of a four-tier diagnostic test to detect students' misconceptions in elementary school science. *Jurnal Pendidikan Sains Indonesia, 9(*3), 367-378. https://doi.org/10.15294/jpsi.v9i3.31900.

Yuberti, Y., Suryani, Y., & Kurniawati, I. (2020). Four-tier diagnostic test with certainty of response index to identify misconception in physics. *Indonesian Journal of Science and Mathematics Education, 3*(2), 245–253. https://doi.org/10.24042/ijsme.v3i2.6061

Yunus, F. W., & Ali, M. (2019). Using multiple-tier diagnostic tests to identify misconceptions in science education. *Journal of Physics: Conference Series, 1157*(2), 022032. https://doi.org/10.1088/1742-6596/1157/2/022032