

The Urgency of Developing a Four-Tier Diagnostic Test Instrument and Its Validity

Hidayatul Aula Hardiyanti¹, Novia Ayu Sekar Pertiwi^{2*}, Ino Angga Putra³, Kartika Wulandari⁴

^{1,2,3,4} Physics Education, KH. A. Wahab Hasbullah University

*Email: novia.as.pertiwi@unwaha.ac.id

ABSTRACT

A valid diagnostic test instrument is an important component in accurately identifying student misconceptions. This study aims to examine the urgency and validity of the Four-Tier diagnostic test instrument on temperature and heat. The method used is research and development (R&D) with the Borg & Gall model adapted from Sugiyono. The validity of the instrument was tested through assessments by material experts and learning evaluation experts using a four-point Likert scale. The results showed that all test items received good to excellent categories, indicating that the instrument has adequate content and construct validity. These findings confirm that a valid Four-Tier diagnostic test instrument can be a reliable tool in uncovering student misconceptions in physics learning.

Keywords: *Diagnostic Tests; Four-Tier; Instrument Validity; Misconceptions; Temperature and Heat.*

INTRODUCTION

Physics learning at the high school level requires a strong mastery of concepts as a foundation for understanding natural phenomena and developing students' scientific reasoning. Physics focuses not only on mastering formulas but also on in-depth conceptual understanding and the ability to logically connect various concepts. However, in practice, many students struggle to develop a sound conceptual understanding, especially in abstract and interrelated topics.

One fundamental problem in physics learning is the persistently high prevalence of student misconceptions. Misconceptions are conceptual understandings that are inconsistent with recognized scientific concepts, but are believed to be true by students, making them difficult to correct through conventional learning. Afriansyah (2022) states that misconceptions can arise from students' erroneous initial understandings, while Suparno (in Irianti, 2021) emphasizes that misconceptions differ from mere lack of understanding because they form distorted cognitive structures. The presence of misconceptions has the potential to hinder the understanding of advanced concepts and reduce the overall quality of physics learning.

The topic of temperature and heat is one of the physics topics that is highly susceptible to misconceptions. Concepts such as the difference between temperature and heat, the mechanisms of heat transfer, and the relationship between temperature and internal energy are often misunderstood by students. Nasution (2024) emphasized that the concepts of temperature and heat, as fundamental components of thermodynamics, are abstract and difficult to observe directly, thus requiring a thorough conceptual understanding. If misconceptions about this material are not identified early, students will struggle to grasp more complex thermodynamic concepts.

Efforts to identify student misconceptions depend heavily on the quality of the assessment instruments used. Unfortunately, learning evaluation in schools is still dominated by conventional tests, such as multiple-choice questions, which only assess whether answers are correct or incorrect without revealing the students' reasoning and level of confidence. These types of tests are unable to distinguish students who truly understand a concept from those who answer correctly by guessing (Permana & Bakri, 2022). As a result, misconceptions often go undetected and persist into subsequent learning.

This situation highlights the urgency of developing diagnostic test instruments capable of uncovering students' conceptual understanding in greater depth. The Four-Tier diagnostic test is one such instrument specifically designed for this purpose. This instrument contains four levels: conceptual answers, level of confidence in the answer, the rationale underlying the answer, and the level of confidence in that rationale (Fariyani & Rusilowati, 2015). This structure allows educators not only to know the final results but also to trace students' thought processes and identify misconceptions more accurately.

However, the development of diagnostic instruments cannot be separated from the aspect of validity. Invalid instruments risk producing an erroneous diagnosis of student understanding. Therefore, validity testing is a crucial stage in the development of a Four-Tier diagnostic test instrument to ensure that each item truly represents the concept being measured and accurately reveals misconceptions. A valid instrument will provide accurate and reliable information as a basis for more effective learning planning.

Based on this description, the development and validity testing of a Four-Tier diagnostic test instrument on the topic of temperature and heat are crucial. A valid and tested instrument is expected to be a reliable diagnostic tool for mapping high school students' misconceptions, providing teachers with a comprehensive picture of students' conceptual understanding before the lesson begins.

METHOD

• Type and Stages of the Research

This study employed a Research and Development (R&D) method by adapting the Borg & Gall model (Sugiyono, 2022). In this article, the research stages were limited to the design validation stage, in accordance with the research objective, which focuses on the urgency of instrument development and the description of the validity test results of a Four-Tier diagnostic test instrument. The research stages implemented consisted of :

- Identification of potential and problems,
- Preliminary data collection,
- Design of the Four-Tier diagnostic test instrument, and
- Expert design validation.

• Research Data

The data used in this study consisted of primary data and secondary data. Primary data were in the form of instrument validity scores provided by physics content experts and educational evaluation experts through a Four-Tier diagnostic test validation sheet. These data were used to determine the level of validity of the developed instrument. Secondary data were obtained from scientific journal articles related to physics misconceptions and the development of Four-Tier diagnostic tests, curriculum documents, and previous studies that served as the basis for developing indicators and test items.

• Data Collection Technique

Data collection was conducted through expert judgment. The developed instrument was evaluated by:

- Physics content experts, who assessed the appropriateness of the content and the scientific correctness of temperature and heat concepts;
- Educational evaluation experts, who assessed item construction, language use, and the appropriateness of the Four-Tier format as a diagnostic instrument.

The assessment was carried out using a validation sheet with a 4-point Likert scale.

• Data Collection Instrument

The data collection instrument was a validation sheet for the Four-Tier diagnostic test instrument. The instrument employed a 4-point Likert scale (1–4) and consisted of 16 validation indicators, which were grouped into three aspects: content appropriateness, item construction, and language and mechanics.

Table 1 Validation Indicators of the Four-Tier Diagnostic Test Instrument

Aspect	Validation Indicators
Content Appropriateness	Test items are aligned with learning indicators and objectives
	The tested material is consistent with indicators, temperature, and heat concepts.
	Answer keys are scientifically correct.

	Incorrect options represent commonly occurring student misconceptions.
	Items are appropriate for senior high school students' cognitive level.
Item Construction	The stem is clear and does not cause ambiguity.
	Answer options are homogeneous in form and content.
	Items do not contain negative statements.
	The item wording does not provide clues to the correct answer.
	Items are arranged in a logical order of difficulty.
	There is a balance between conceptual and computational questions.
Language and Mechanics	Language follows Indonesian language conventions.
	Physics terminology is used accurately and consistently.
	Language is communicative and easy for students to understand
	No ambiguous terms or excessively long sentences are used.
	Symbols, units, and physics notation follow SI standards.

These indicators were used by validators to systematically assess each test item.

- Data Analysis Technique

Data analysis focused on the validity of the Four-Tier diagnostic test instrument. The evaluation was conducted by physics content experts and educational evaluation experts using a 4-point Likert scale with categories 4 = very valid, 3 = valid, 2 = less valid, and 1 = not valid. The validity scores were then interpreted based on the instrument validity categories presented in the following table.

Table 2. Validity Categories of the Four-Tier Diagnostic Test Instrument

Total Score	Category	Description
$48 \leq S \leq 64$	Very Good	Valid for use without revision
$32 \leq S < 48$	Good	Valid for use with minor revisions
$16 \leq S < 32$	Fair	Requires significant revision before use
$0 \leq S < 16$	Poor	Validity is not recommended for use

RESULT AND DISCUSSION

Result

The validity of the Four-Tier diagnostic test instrument on temperature and heat was evaluated through expert judgment involving physics content experts and educational evaluation experts. The validation focused on three main aspects, namely content appropriateness, item construction, and language and mechanics, as described in the research method.

The results of expert validation indicate that the average validity scores for all test items ranged from 48 to 55. Based on the predetermined validity criteria, most items were classified in the Very Good category, while a small number of items were categorized as Good. None of the items fell into the Fair or Poor categories. These results demonstrate that, overall, the developed Four-Tier diagnostic test instrument meets the established validity standards.

The consistent achievement of high validity scores across items suggests that the instrument has strong content validity. The physics content experts confirmed that the concepts assessed in the test items were scientifically accurate and aligned with the key concepts of temperature and heat taught at the senior high school level. In addition, the distractors provided in Tier 1 and Tier 3 were considered representative of common student misconceptions, which is a critical characteristic of a diagnostic instrument.

From the perspective of construct validity, the educational evaluation experts indicated that the Four-Tier structure was appropriately designed to capture not only students' answers but also their reasoning and confidence levels. The logical relationship between Tier 1 (answer choice), Tier 2 (confidence level), Tier 3 (reasoning), and Tier 4 (confidence in reasoning) was assessed as clear and coherent, enabling a more comprehensive diagnosis of students' conceptual understanding.

Discussion

The results of the validity test highlight the urgency and relevance of developing a Four-Tier diagnostic test instrument for the topic of temperature and heat. Physics concepts in this domain are well known for being abstract and prone to misconceptions, such as confusion between temperature and heat or misunderstandings of heat transfer mechanisms. Conventional assessment instruments, which typically emphasize final answers, are often insufficient to detect these misconceptions accurately.

The high validity scores obtained in this study indicate that the developed instrument is capable of addressing these limitations. The strong content validity ensures that each test item measures the intended physics concepts accurately, while the construct validity of the Four-Tier format enables deeper exploration of students' conceptual reasoning and confidence. This finding supports previous studies suggesting that Four-Tier diagnostic tests provide richer diagnostic information than traditional multiple-choice tests (Caleon & Subramaniam, 2010).

Expert feedback during the validation process primarily concerned minor revisions related to wording clarity, refinement of indicators, and alignment between conceptual questions and reasoning options. After incorporating these suggestions, the revised instrument demonstrated improved quality, reflected in the predominance of the Very Good validity category. This result emphasizes the importance of expert validation as a critical step in ensuring the quality and feasibility of educational assessment instruments (Arikunto, 2018).

Furthermore, the validity results reinforce the argument that developing a diagnostic instrument is not merely a technical task but a pedagogical necessity. A valid Four-Tier diagnostic test enables teachers to identify misconceptions more precisely, which is essential for designing effective remedial instruction. In this sense, the development of a valid diagnostic instrument on temperature and heat is urgent, particularly in senior high school physics, where conceptual understanding forms the foundation for more advanced topics.

In summary, the results demonstrate that the developed Four-Tier diagnostic test instrument possesses strong validity and is suitable for use as a diagnostic tool. The findings also underline the critical role of validity testing in the development process, ensuring that the instrument can function effectively in identifying students' conceptual understanding and misconceptions in physics learning.

CONCLUSIONS

The development of a Four-Tier diagnostic test instrument for temperature and heat topics is pedagogically necessary to address the limitations of conventional assessments in diagnosing students' conceptual understanding. The results of expert validation indicate that the developed instrument meets the established validity criteria, with most items categorized as Very Good and the remaining items categorized as Good.

These findings demonstrate that the instrument has strong content validity, appropriate item construction, and clear language, making it suitable for use as a diagnostic assessment tool. Therefore, the Four-Tier diagnostic test instrument developed in this study is considered valid and feasible for identifying students' understanding and misconceptions on temperature and heat concepts at the senior high school level.

REFERENCES

- Arikunto, S. (2018). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Afriansyah, E. A. (2022). Peran RME terhadap Miskonsepsi Siswa MTs pada Materi Bangun Datar Segi Empat. *Mosharafa: Jurnal Pendidikan Matematika*, 11(3), 359–368. <https://doi.org/10.31980/mosharafa.v11i3.2102>
- Caleon, I. S., & Subramaniam, R. (2010). 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. *Research in Science Education*, 40(3), 313–337. <https://doi.org/10.1007/s11165-009-9122-4>
- Fariyani, Q., & Rusilowati, A. (2015). Pengembangan tes diagnostik four-tier untuk mengungkap miskonsepsi fisika siswa SMA. *Jurnal Pendidikan Fisika Indonesia*, 11(1), 41–49.
- Irianti, E. (2021). Identifikasi Miskonsepsi Siswa pada Materi Cahaya Menggunakan Four Tier Diagnostic Test. *Jurnal Pendidikan Fisika Undiksha*, 11(2), 1-10.
- Nasution, S. K. (2024). Analisis Kesulitan Peserta Didik Dalam Memahami Konsep Fisika Pada Materi Termodinamika Di SMA. *ejournal.mtpublisher*, Vol. 3, 11.
- Permana, H., & Bakri, F. (2022). Pelatihan Pembuatan Instrumen Tes Diagnostik Fisika Untuk

- Mengetahui Miskonsepsi Pada Peserta Didik. *Mitra Teras: Jurnal Terapan Pengabdian Masyarakat*, 1(1), 29–36. <https://doi.org/10.58797/teras.0101.05>
- Rofiudin, M. (2022). Pengembangan Tes E-Diagnostik Empat Tingkat Berbasis Web Untuk Mengungkap Miskonsepsi Mahasiswa Calon Guru Fisika Pada Materi Termodinamika [Thesis]. Universitas Islam Negeri Walisongo Semarang.
- Sugiyono. (2022). *Metode penelitian dan pengembangan (Research and Development)*. Bandung: Alfabeta.