P-ISSN: 2723-8156
e-ISSN: 2723-8164
JUNE 2026

U-Teach: Journal Education of Young Physics Teacher 7 (1) (2026) 77-85
https://jsr.unuha.ac.id/index.php/U-Teach

Cognitive Achievement Paradox in High School Physics: Discrepancy Between Test Performance and Conceptual Understanding

Hermansyah[1*], Syarif Fitriyanto[1], Sri Hartini[1], Cici Aprina[1], Armansyah Putra[1]

[1]Department of Physics Education, Universitas Samawa, Sumbawa, Indonesia
**Corresponding author*: hermansyah.fis92@gmail.com

Article History:
Received: Mei 10, 2026
Revised: Mei 17, 2026
Accepted: Juni 03, 2026
Published: Juni 13, 2026

Keywords: *Cognitive achievement paradox, conceptual understanding, heat and temperature, physics education, rote learning*

Abstract: *This study investigated the Cognitive Achievement Paradox in high school physics learning under conditions where high grades are accompanied by shallow conceptual understanding. A convergent-parallel mixed-methods design involved 48 11th-grade students at a public high school in Sumbawa. Data were collected through a concept test on Heat and Temperature, structured interviews, and classroom observations. Quantitative analysis using SPSS 26 showed a class mean of 89.81 (SD = 8.74), 95.8% achieved the KKM (≥75), and Shapiro–Wilk showed non-normality (W = 0.897, p < .001). However, qualitative findings revealed an inability to contextualize thermal phenomena, a reliance on memorization, and inadequate integration of simulation media. Results confirmed that calculation-based assessment systematically overestimates true understanding, threatening construct validity and scientific literacy. A quasi-experimental study testing the Simulation-Anchored Contextual Assessment model based on PhET simulations as an embedded formative assessment tool is recommended.*

Abstrak: Penelitian ini menginvestigasi Cognitive Achievement Paradox pada pembelajaran fisika SMA dengan kondisi di mana nilai tinggi disertai pemahaman konseptual yang dangkal. Desain mixed-methods konvergen-paralel melibatkan 48 siswa Kelas XI di SMA negeri Sumbawa. Data dikumpulkan lewat tes konsep tentang Panas dan Suhu, wawancara terstruktur, dan observasi kelas. Analisis kuantitatif menggunakan SPSS 26 yang menunjukkan rata-rata kelas 89,81 (SD=8,74), 95,8% mencapai KKM (≥75), dan Shapiro–Wilk menunjukkan non-normalitas (W=0,897, p<.001). Namun temuan kualitatif mengungkap ketidakmampuan mengontekstualkan fenomena termal, ketergantungan pada hafalan, dan integrasi media simulasi yang kurang. Hasil menegaskan bahwa penilaian berbasis perhitungan secara sistematis melebihkan pemahaman sejati, mengancam validitas konstrak dan literasi sains. Disarankan penelitian kuasi-eksperimental menguji model Simulation-Anchored Contextual Assessment berbasis simulasi PhET sebagai alat penilaian formatif tertanam.

INTRODUCTION

Physics education research has long grappled with a persistent tension between measurable academic performance and authentic conceptual understanding (Docktor & José, 2014; Hake, 1998). Students may achieve high scores on standardized assessments while simultaneously harboring significant misconceptions or possessing only surface-level procedural knowledge that collapses when confronted with novel, contextualized problems. This phenomenon where quantitative achievement metrics diverge substantially from genuine conceptual comprehension constitutes what the present study terms the Cognitive Achievement Paradox.

In Indonesian secondary education, physics consistently ranks among subjects with the highest affective barriers, with students frequently reporting anxiety toward mathematical calculations while simultaneously expressing curiosity about natural phenomena (Hajar, 2025). The national examination system and minimum competency criteria (KKM) create institutional incentives for score attainment, potentially at the expense of deeper scientific reasoning. As Biggs & Tang (2011) argue, educational systems that overemphasize summative grading tend to inadvertently promote surface learning strategies characterized by rote memorization and formula application over the deep learning approaches that foster transferable conceptual understanding.

The novelty of this study lies in its systematic empirical documentation of the Cognitive Achievement Paradox specifically within the high school physics context in Sumbawa, a setting that remains significantly underrepresented in the international physics education research literature (Mullis et al., 2021). While previous investigations have examined similar paradoxes in Western higher education (Guyottot & Thelisson, 2026) and East Asian secondary school systems (Green & Phan, 2025), the unique configuration of exam-driven KKM accountability, resource constraints, and underutilization of simulation in schools creates a distinct pedagogical ecology that requires dedicated empirical scrutiny. This study provides the first mixed-methods characterization of this paradox in the regional context of Sumbawa, offering locally grounded evidence to inform national assessment reform discourse.

The implications of this paradox extend beyond individual student outcomes. At the systemic level, inflated test scores can mask significant deficiencies in scientific literacy, mislead educators about the effectiveness of their instructional approaches, and generate misplaced confidence in curriculum adequacy (Deslauriers et al., 2019). Within the specific context of thermodynamics, the domain of Heat and Temperature examined in this study conceptual understanding demands that students bridge abstract thermodynamic principles with observable macroscopic phenomena, a cognitive requirement that fundamentally exceeds the ability to recall and apply formulae.

Previous research in physics education has documented the inadequacy of traditional test scores as proxies for conceptual understanding. Diagnostic instruments such as the Force Concept Inventory have consistently demonstrated that high-performing students, as measured by conventional examinations, may simultaneously hold deeply entrenched alternative conceptions (Docktor & José, 2014). However, the majority of such investigations have been conducted in Western educational contexts, with comparatively limited exploration of this paradox in Southeast Asian secondary schools, where examination-driven cultures and structural constraints create distinct pedagogical dynamics.

This study addresses three research questions: (1) What does the distribution of physics test scores reveal about student achievement in a thermodynamics unit? (2) To what extent do quantitative test results align with qualitative evidence of conceptual understanding from interviews and classroom observation? (3) What pedagogical and assessment implications arise from the identified discrepancy?

THEORETICAL FRAMEWORK

Conceptual Understanding versus Procedural Knowledge

The distinction between conceptual and procedural knowledge in science education has been extensively theorized (Chi & Wylie, 2014; Docktor & José, 2014). Conceptual knowledge refers to the organized network of interrelated principles, phenomena, and causal relationships within a domain, enabling flexible application across varied contexts. Procedural knowledge, by contrast, encompasses the step-by-step algorithms and formulaic operations that can be executed without reference to underlying conceptual structures. In physics, this distinction manifests prominently: a student may correctly apply the formula $Q=mc\Delta T$ without understanding the molecular basis of heat transfer or solve Black's principle (*Azas Black*) problems algorithmically without comprehending their physical significance. ICAP framework posits that authentic learning the kind that builds transferable conceptual structures requires active cognitive engagement that moves beyond passive reception and memorization toward constructive and interactive processing (Chi & Wylie, 2014).

Surface Learning and Rote Memorization

Rote learning, an extreme form of surface processing, is particularly prevalent in high-stakes examination environments where the predictability of test formats enables students to achieve high scores through memorization alone (Biggs & Tang, 2011). This creates the paradox examined in the present study: students who have successfully memorized thermodynamic formulae and standard problem types achieve high KKM scores, yet are unable to explain why a metal spoon becomes hot when placed in boiling water or why gaps are incorporated in bridge construction phenomena directly explicable by the same principles they ostensibly understand. Marton & Säljö (1976, as cited in Biggs & Tang, 2011) foundationally distinguished surface from deep approaches to learning, demonstrating that assessment design profoundly shapes the cognitive strategies learners adopt.

Assessment Validity and Scientific Literacy

The validity of an assessment instrument refers to the degree to which it measures what it purports to measure (Docktor & José, 2014). When test instruments are dominated by closed-form calculation problems with predictable solution pathways, they may effectively measure procedural fluency while failing to assess conceptual comprehension. Deslauriers et al., (2019), in a landmark PNAS study, demonstrated that students who felt they had learned more through passive instruction performed better on immediate tests designed around familiar problem types, while students taught through active learning demonstrated superior performance on conceptually richer assessments highlighting the profound sensitivity of outcome measures to assessment design.

This assessment validity problem is compounded in Indonesian secondary physics education by the structure of KKM-based evaluation, where passing a predetermined threshold becomes the operative objective, potentially decoupling instructional goals from genuine scientific literacy development (Rusilowati et al., 2016). The PISA framework for scientific literacy explicitly requires that students be able to apply scientific knowledge to real-world contexts, explain phenomena scientifically, and evaluate scientific evidence capabilities that standard formula-application tests do not assess (OCED, 2023).

RESEARCH METHOD

Research Design

This study employed a convergent parallel mixed-methods design (Creswell, 2009), integrating quantitative analysis of test score data with qualitative evidence from structured interviews and systematic classroom observation. The concurrent collection and subsequent integration of both data streams enabled triangulation of findings and a comprehensive characterization of the achievement-understanding discrepancy.

Participants

Participants comprised 48 students (N=48) from Class XI of a state senior high school in Sumbawa, West Nusa Tenggara, Indonesia. Data were collected on April 2026, during a regular Physics class session on Heat and Temperature, taught by a certified physics teacher with over ten years of instructional experience. Four students two high-achieving and two mid-achieving based on prior academic records, were purposively selected for in-depth interviews to ensure variation in perspective.

Instruments

(1) Physics Concept Understanding Test: A written examination covering key concepts in Heat and Temperature, including the concept and definition of temperature, temperature scale conversions (Celsius, Fahrenheit, Kelvin, and Réaumur), thermal expansion, specific heat capacity, latent heat, and changes of state. Scores were recorded on a 0–100 scale, compared against the school's KKM threshold of 75.

(2) Structured Interview Guide: Semi-structured protocols for both teacher and student interviews. Teacher interviews explored perceived student difficulties, instructional strategies, and resource availability. Student interviews probed learning preferences, self-reported strategies, and ability to articulate phenomenological explanations of thermal concepts.

(3) Classroom Observation Sheet: A systematic instrument used to evaluate instructional delivery quality, media utilization, student engagement, and the integration of experimental or simulation-based activities, based on established criteria for quality physics instruction (Docktor & José, 2014).

Data Analysis

Quantitative data were analyzed using descriptive statistics (mean, median, mode, standard deviation, skewness, kurtosis, quartiles) computed with SPSS version 26.0. Normality was assessed using the Shapiro-Wilk test, which is recommended for samples of $n < 50$ due to its superior statistical power compared to the Kolmogorov-Smirnov test (Razali & Wah, 2011). Qualitative data from interviews and observations were analyzed thematically through iterative coding, with codes organized into overarching themes aligned with the research questions. Integration of quantitative and qualitative findings followed a joint display approach (Creswell & Plano Clark, 2018) to enable direct comparative analysis of both data streams.

RESULTS AND DISCUSSION

Quantitative Analysis: Test Score Distribution

Descriptive statistics for the 48 student test scores are presented in Table 1.

Table 1. Descriptive Statistics of Concept Understanding Test Scores

Statistic	Value
Mean	89.81
Median	90.00
Mode	80 (appears 11 times)
Standard Deviation	8.74
Variance	76.37
Minimum Score	67
Maximum Score	100
Q1 / Q3	80.00 / 98.00
IQR	18.00
Skewness	−0.42 (negatively skewed)
Kurtosis (excess)	−0.78 (platykurtic)
KKM Pass Rate (≥75)	95.8% (46 of 48 students)
Shapiro-Wilk (W, p)	W = 0.897, p = 0.0005 (non-normal)

The class mean of 89.81 places overall performance within the "very high" category according to standard Indonesian scoring rubrics, and 95.8% of students (46 of 48) achieved scores at or above the KKM threshold. A total of 27 students

(56.3%) scored 90 or above; 19 students (39.6%) scored between 80 and 89; only one student scored in the 70–79 range, and one scored below 70. However, the Shapiro-Wilk test (W=0.897, p=0.0005) rejected the null hypothesis of normality, confirming the distribution is non-normal.

The distribution exhibited negative skewness (−0.42), indicating score concentration at the upper extreme, and platykurtic spread (excess kurtosis = −0.78), reflecting a bimodal tendency with pronounced clustering at both 80 (n = 11, 22.9%) and the 98–100 range (n = 16, 33.3%). This bimodality is diagnostically significant: it suggests the existence of two qualitatively distinct learner subgroups those who have mastered basic formula application and those who have developed more fluent pattern-matching for standard problem variants without necessarily developing deeper conceptual understanding in either case.

Qualitative Analysis: Interview and Observation Findings

Theme 1: Contextualisation Barrier. Across all four student interviews, participants were unable to spontaneously generate real-world examples of thermal phenomena beyond those explicitly taught in class. When prompted to explain the thermal physics of a traditional cooking implement or the engineering rationale for expansion gaps in bridge construction applications directly explainable by Heat and Temperature principles students exhibited confusion and resorted to restating formulae without explanatory meaning.

Theme 2: Rote Learning as the Dominant Epistemic Strategy. All four students and the teacher independently confirmed that examination preparation dominated learning activity, with students focusing intensively on memorizing formula variants and practicing standard calculation templates. Student engagement was highest during formula-application exercises but dropped markedly during phenomenological discussion. The teacher explicitly acknowledged time pressure as a structural constraint that prevented exploratory, context-based instruction.

Theme 3: Underutilisation of Simulation and Interactive Media. Classroom observation revealed that while the teacher employed Discovery Learning as a nominal instructional framework and initiated productive introductory discussion linking thermodynamics to everyday life, the core instructional sequence relied predominantly on verbal explanation and worked examples. No simulation software (e.g., PhET Interactive Simulations), laboratory demonstration, or interactive digital media was employed during the observed session, despite the school having adequate technological infrastructure.

Integration of Quantitative and Qualitative Findings

The joint display in Table 2 synthesizes the central paradox of this study.

Table 2. Joint Display: Quantitative Performance vs. Qualitative Understanding Evidence

Dimension	Quantitative Evidence	Qualitative Evidence
Achievement level	Mean = 89.81; 95.8% pass KKM	Students cannot explain everyday thermal phenomena
Learning strategy	-	Rote memorization; formula recall dominant
Conceptual depth	High scores suggest mastery	Confusion on contextual / phenomenological tasks
Instructional media	-	No simulation or interactive media used
Assessment type	Calculation-dominant test	Contextualization not assessed

The findings of this study provide empirical illustration of the Cognitive Achievement Paradox in a naturalistic Indonesian secondary physics context. The contradiction between a class mean of 89.81 and pervasive rote dependence with contextualization failure is not an anomaly but a predictable consequence of assessment-driven instruction.

This paradox aligns with patterns documented in the physics education research literature. Docktor & José (2014) note that traditional physics examinations dominated by algorithmic problem-solving correlate poorly with conceptual diagnostic measures. Deslauriers et al., (2019) experimental evidence demonstrated that conventional test performance can be systematically decoupled from actual conceptual learning gains, particularly when assessment instruments favor procedural familiarity. The present study extends this evidence base to secondary physics education in an Indonesian context, where examination pressure appears to amplify the paradox beyond what has been reported in Western settings.

The bimodal score distribution warrants particular interpretive attention. Rather than indicating natural classroom heterogeneity, the clustering at 80 and 98–100 more plausibly reflects two categories of procedural fluency: students who have memorized sufficient formula templates to solve standard problems reliably, and students who have additionally developed efficient pattern recognition for multiple problem variants. Critically, neither group necessarily possesses the deep conceptual structures required for generative reasoning about novel thermodynamic scenarios a conclusion directly supported by the interview evidence. This pattern corroborates findings from Kapul et al., (2023), who documented that students frequently conflate temperature and heat as equivalent concepts and misapply thermal expansion principles, demonstrating that high procedural scores do not preclude persistent conceptual errors on the same material.

The underutilisation of simulation tools such as PhET Interactive Simulations represents a significant missed instructional opportunity. Research consistently demonstrates that well-designed virtual environments scaffold the construction of phenomenological mental models that purely symbolic instruction cannot provide (Ebrahimi & Ramaprasad, 2025; Lehrman, 2025; Wieman et al., 2010). The absence of simulation-based activity during instruction on abstract processes such as temperature scale relationships, thermal expansion, and changes of state deprived students of the visual-experiential grounding needed to connect symbolic representations to physical reality. This finding is directly corroborated by Nurulfajri et al., (2024), whose literature review of PhET integration in Heat and Temperature instruction specifically found that simulation-based learning significantly improved conceptual understanding scores, with the greatest gains occurring in precisely those sub-topics thermal expansion and changes of state where the present study's qualitative data revealed the deepest contextualisation failures. Furthermore, Jamal et al., (2025) demonstrated that structured Problem-Based Learning integrated with digital tools produced measurable gains in Heat and Temperature concept mastery in an Indonesian Senior High School context comparable to this study reinforcing the inference that the absence of such approaches in the observed classroom constitutes a substantive instructional deficit rather than a neutral pedagogical choice.

These findings carry direct implications for the construct validity of physics assessment in Indonesian secondary schools. When a test dominated by algorithmic calculation items yields near-perfect class performance while interview evidence reveals an inability to explain the underlying phenomena, the test is demonstrably not assessing conceptual understanding as its intended construct (Docktor & José, 2014). Reforming assessment to include phenomenological explanation tasks, visual interpretation items, and real-world contextualisation problems consistent with PISA scientific literacy frameworks (OCED, 2023) would more accurately reveal the actual distribution of conceptual understanding. The comparison of Indonesian and Irish physics teaching practices documented by Sudirman et al., (2023) in Frontiers in Education is instructive in this regard: Indonesian physics teachers were found to rely significantly more on direct instruction and examination preparation than their counterparts, a structural pattern that the present study's teacher interview evidence confirms and its score-understanding discrepancy quantifies.

The teacher's acknowledgment of time constraints as a barrier to experiential instruction illuminates a broader systemic tension: curriculum coverage and KKM achievement demands consume the instructional time that might otherwise support deeper, discussion-based exploration of phenomena. This structural constraint is not unique to Indonesia but is amplified in systems where high-stakes summative assessment carries significant institutional weight (Biggs & Tang, 2011; Manasikana et al., 2026). A systematic review of physics learning media in Indonesian senior high schools from 2014–2024 by Anandita et al., (2025) confirms that while digital and simulation-based media have been increasingly adopted at the national level, implementation remains uneven, with schools in eastern Indonesia the regional context of the present study showing the widest gap between media availability and actual instructional integration. Addressing the Cognitive Achievement Paradox therefore requires systemic-level reform beyond individual instructional change, encompassing assessment redesign, curriculum depressurisation, and sustained investment in pedagogical capacity for simulation-integrated and context-based teaching.

CONCLUSION

This study has documented and analyzed the Cognitive Achievement Paradox in a high school physics classroom, demonstrating through mixed-methods evidence that high test scores on a Heat and Temperature concept assessment did not correspond to genuine conceptual understanding among 48 Class XI students. The quantitative data revealed strong aggregate performance (mean = 89.81; KKM pass rate = 95.8%), while qualitative data from interviews and observation revealed pervasive rote learning, an inability to contextualize thermal phenomena, and the absence of meaningful simulation-based instruction.

Three principal conclusions are drawn. First, standard calculation-based physics tests, as currently implemented, are insufficient proxies for conceptual understanding and systematically overestimate actual student comprehension. Second, the bimodal score-distribution reflects differentiated procedural fluency rather than genuine conceptual heterogeneity. Third, structural factors time pressure, KKM-centered instruction, and inadequate utilization of simulation media collectively sustain the paradox by reinforcing surface learning strategies.

These findings have direct implications for assessment reform, pedagogical innovation, and the design of subsequent intervention research. Future research should investigate the Simulation-Anchored Contextual Assessment model using PhET simulations as embedded formative assessment tools to test whether model and simulation-based pedagogical approaches can disrupt the rote learning cycle and produce genuine improvements in both test performance and conceptual understanding.

REFERENCES

Anandita, A. S., Negoro, R. A., & Rahmawati, L. (2025). Transforming physics learning media in indonesian senior high schools: a decade of systematic innovation review (2014-2024). *Physics Education Journal*, *8*(1).

Biggs, J., & Tang, C. (2011). Teaching for quality learning at university. What the student does (4th Edn.). In *Innovations in Education and Teaching International* (Vol. 50, Number 4).

Chi, M. T. H., & Wylie, R. (2014). The icap framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4). https://doi.org/10.1080/00461520.2014.965823

Creswell, J. W. (2009). Research design: qualitative, quantitative, and mixed methods approaches. *Intercultural Education*, *3*(2). https://doi.org/10.1080/14675980902922143

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research. In *Organizational Research Methods* (Vol. 12, Number 4).

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(39). https://doi.org/10.1073/pnas.1821936116

Docktor, L., & José, P. M. (2014). Synthesis of discipline-based education research in physics. *Physical Review Physics Education Research*, *10*.

Ebrahimi, A., & Ramaprasad, H. (2025). Interaction design strategies for socio-spatial embodiment in virtual world learning. *Virtual Worlds*, *4*(3). https://doi.org/10.3390/virtualworlds4030030

Green, A., & Phan, H. (2025). Culture, context and East Asian high-performing education systems. *Research Papers in Education*. https://doi.org/10.1080/02671522.2025.2567396

Guyottot, O., & Thelisson, A. S. (2026). Navigating the paradoxical currents of higher education: Triggers of tensions for business school top managers. *Higher Education Policy*. https://doi.org/10.1057/s41307-025-00437-9

Hajar, S. (2025). Interpretasi fenomenologis tentang pengalaman siswa dalam penanggulangan math anxiety. *Jurnal Ilmiah IPA Dan Matematika (JIIM)*, *3*(1).

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1). https://doi.org/10.1119/1.18809

Jamal, S. N., Ibrahim, N. H., Abd Halim, N. D., Surif, J., Saputra, A., Tania, L., & Sunyono, S. (2025). Problem-based learning integrated with the direct creative thinking process (pbl-dct) module in enhancing creativity in chemistry. *Journal of Chemical Education*. https://doi.org/10.1021/acs.jchemed.3c01237

Lehrman, A. L. (2025). Embodied learning through immersive virtual reality: Theoretical perspectives for art and design education. *Behavioral*

Sciences, *15*(7). https://doi.org/10.3390/bs15070917

M. Kapul, V. Lantik, & K.A. Astiti. (2023). Analisis miskonsepsi siswa dan alternatif remediasinya pada konsep suhu dan kalor. *Jurnal Pendidikan Dan Pembelajaran IPA Indonesia*, *13*(1). https://doi.org/10.23887/jppii.v13i1.56275

Manasikana, A., Harani, D. A., Nabila, M., Azyati, N., Mumtaza, U., & Sukma, Z. (2026). Telaah literatur tentang efektivitas evaluasi sumatif dalam menilai pencapaian kompetensi peserta didik. *An-nahdloh: Journal of Education and Islamic Studies*, *1*(2).

Mullis, Martin, & Davier, V. (2021). Timss 2023 assessment framework. *TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.*

Nurulfajri, H., Lestari, I. F., & Warliani, R. (2024). Optimalisasi Pembelajaran Fisika dengan PhET Simulation: Studi Kuantitatif pada Siswa SMA. *Pendidikan Dan Ilmu Fisika*, *4*(2). https://doi.org/10.52434/jpif.v4i2.42124

OCED. (2023). PISA 2022 Results (Volume I): The State of Learning and Equity in Education. *OECD Publishing*, *I*(volume I).

Rusilowati, A., Nugroho, S. E., & Susilowati, S. M. (2016). Development of science textbook based on scientific literacy for secondary school. *Jurnal Pendidikan Fisika Indonesia*, *12*(2). https://doi.org/10.15294/jpfi.v12i2.4252

Sudirman, S., Kennedy, D., & Soeharto, S. (2023). The teaching of physics at upper secondary school level: A comparative study between Indonesia and Ireland. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1118873

Wieman, C. E., Adams, W. K., Loeblein, P., & Perkins, K. K. (2010). Teaching physics using phet simulations. *The Physics Teacher*, *48*(4). https://doi.org/10.1119/1.3361987