



Development of a Distractor-Based Cognitive Diagnostic Test (DB-CDT) on Thermodynamics Material to Detect High School Students' Data Literacy Skills

Supahar^{1*}, Ayu Mar'ati Barokatun Ni'mah¹, Reny Afifah¹, Putri Edryna Larasati¹

¹Physics Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia.

Received: December 15, 2025

Revised: January 28, 2026

Accepted: February 25, 2026

Published: February 28, 2026

Corresponding Author:

Supahar

supahar@uny.ac.id

DOI: [10.29303/jppipa.v12i2.13899](https://doi.org/10.29303/jppipa.v12i2.13899)

 Open Access

© 2026 The Authors. This article is distributed under a (CC-BY License)



Abstract: This study addresses the lack of specialized diagnostic instruments capable of pinpointing specific cognitive failures in physics data literacy, particularly within Thermodynamics. To bridge this gap, a Distractor-Based Cognitive Diagnostic Test (DB-CDT) was developed and validated for eleventh-grade high school students in Yogyakarta. Utilizing a Research and Development (R&D) approach with the ADDIE framework, the instrument was built upon the DINA (Deterministic Input, Noisy-AND Gate) cognitive diagnostic model. Expert validation by three specialists yielded an excellent average V-Aiken value of 1.00. Empirical testing on a sample of 540 students across high, medium, and low-tier schools confirmed high psychometric quality: an average INFIT MNSQ of 1.00 (within the 0.77–1.30 range), item difficulty between -2 and +2, and high reliability (0.92). DB-CDT distinguishes itself by using a Q-Matrix to link specific distractors to cognitive errors, such as failures in sign conversion or process interpretation. Analysis of student profiles revealed that the experimental group—utilizing targeted diagnostic feedback—achieved superior mastery in analyzing (96.6%) and interpreting data (90%), whereas the control group, receiving conventional assessment, struggled significantly with inference and synthesis. Theoretically, this research advances the application of Cognitive Diagnostic Models in physics education. Practically, it provides educators with a sensitive tool to profile individual student weaknesses, allowing for more focused instructional interventions in complex scientific topics.

Keywords: Data literacy; Distractor-based cognitive diagnostic test; Thermodynamics

Introduction

Modern science education increasingly emphasizes the mastery of 21st-century skills, where data literacy—encompassing the complete cycle of collecting, managing, critically evaluating, and communicating information—serves as a fundamental competency for understanding complex scientific phenomena (OECD, 2023). The urgency of equipping students with this capability is further reinforced by the rapid advancement of information technology and the prevalence of big data in the modern era (Lee et al., 2024; Mandinach et al., 2021). In the specific context of Physics,

mastering data literacy is vital as it allows students to transform experimental observations into a solid conceptual understanding, moving beyond simple formula memorization toward deeper analytical proficiency.

The transition from a global imperative to specific disciplinary urgency is most evident in Thermodynamics, a subject that often presents significant barriers for high school students due to its abstract and multidimensional nature (Natalis et al., 2025; Prasanth et al., 2023). Core concepts such as internal energy, heat, and entropy require high-level logical reasoning and the fluid ability to transition between verbal, mathematical, and graphical

How to Cite:

Supahar, Ni'mah, A. M. B., Afifah, R., & Larasati, P. E. Development of a Distractor-Based Cognitive Diagnostic Test (DB-CDT) on Thermodynamics Material to Detect High School Students' Data Literacy Skills. *Jurnal Penelitian Pendidikan IPA*, 12(2), 80-89. <https://doi.org/10.29303/jppipa.v12i2.13899>

representations. Because processes like the Carnot cycle and engine efficiency are inherently visualized through quantitative data—specifically P-V curves and tabular forms—students' success depends heavily on their capacity to read and interpret these visualizations (Guglietti, 2023; Munandar et al., 2024; Suryadi et al., 2021). Ultimately, the failure to link these data representations with underlying Physics principles is often the root cause of persistent misconceptions, making data literacy an essential prerequisite for success in learning Thermodynamics.

Traditional assessment instruments, particularly those relying on total-score multiple-choice formats, fail to provide the actionable cognitive information necessary for effective instructional remediation. The fundamental problem is that these instruments prioritize aggregate scores, which only indicate a general level of mastery while masking the specific nature of a student's cognitive gaps (Scaria et al., 2025). Without a granular attribute diagnosis, teacher interventions are inevitably poorly targeted, as they lack the data needed to address the root causes of student errors (Pusporini et al., 2026). In the context of Thermodynamics, this lack of diagnostic depth is particularly detrimental; a low score fails to reveal whether a student's struggle stems from a basic inability to read axes, a failure to identify complex trends, or a fundamental misconception in differentiating between isothermal and adiabatic process graphs (Laliyo et al., 2020; Rupp et al., 2010). Consequently, when assessments provide only surface-level data rather than a detailed cognitive map, deep-seated misconceptions persist unaddressed, hindering the student's progression toward mastering complex physical systems.

The Cognitive Diagnostic Test (CDT) emerges as a superior assessment paradigm due to its ability to map student responses onto a profile of specific cognitive attribute mastery. Unlike Item Response Theory (IRT), which focuses on general latent ability, CDT utilizes Cognitive Diagnostic Models (CDM) to identify the set of cognitive attributes (Q-matrix) required to answer each item (Davison et al., 2023; Lian et al., 2023; Net, 2023). Therefore, the application of CDT is ideal for dissecting the cognitive structure of students concerning their data literacy skills in Thermodynamics.

Although CDT has been widely developed in the field of science, there is a significant research gap concerning instruments that explicitly and focused diagnose the specific sub-skills of Data Literacy within the context of Thermodynamics. Existing CDT research often focuses on conceptual misconceptions or general mathematical attributes, rather than on the critical attributes related to the interpretation of visual data (P-V, T-S graphs), which is characteristic of Thermodynamics (Li, 2023). This gap indicates the need

for the development of an instrument specifically designed to map the crucial dimensions of physics data literacy.

To address this gap, this study proposes the development of a Distractor-Based Cognitive Diagnostic Test (DB-CDT), which is a methodological enhancement of the standard CDT. In the DB-CDT approach, each distractor is meticulously designed to reflect error patterns or misconceptions associated with the failure to master a specific Data Literacy attribute (Bulut et al., 2024; Mulyani et al., 2021; Soeharto, 2021). By analyzing the incorrect response choices (diagnostic distractors), DB-CDT is able to provide significantly richer and more precise information about the reason why students fail to master a particular attribute, and not just whether they failed (Desstya et al., 2025; Laliyo et al., 2022).

The main novelty of this research lies in the integration and implementation of the DB-CDT model to specifically detect Data Literacy attributes within the domain of Thermodynamics. Few studies have methodologically combined the CDT framework with detailed distractor analysis for a hard skill domain such as the interpretation of graphical data in Physics. The outcome of this development is an instrument with high diagnostic accuracy, which can significantly contribute to the development of science education assessment theory, particularly in the measurement of complex abilities.

Therefore, the main objectives of this study are to develop an effective DB-CDT instrument to diagnose the data literacy profile of high school students on Thermodynamics material, to determine the feasibility of the Distractor-Based Cognitive Diagnostic Test (DB-CDT) instrument for diagnosing high school students' data literacy profiles, and to diagnose students' response attributes from the data literacy profile. The practical contribution of this research is the availability of a reliable assessment tool for teachers and researchers, whose diagnosis can be used as a basis for designing personalized remedial learning interventions, thereby fundamentally enhancing students' mastery of Thermodynamics concepts and data literacy competence (Mandinach et al., 2021).

Method

Time and Place of Research

The research was conducted during the odd semester of the 2025/2026 academic year. The study was carried out at six different High Schools (SMA) in Yogyakarta, specifically selected to represent high, medium, and low academic achievement tiers.

Research Design

This study utilizes a Research and Development (R&D) approach following the ADDIE framework (Branch, 2009). The research population encompasses eleventh-grade Physics students across Yogyakarta, with a sample size of 540 students selected through a purposive sampling technique. This technique was employed to ensure that the selected respondents had the prerequisite knowledge in Thermodynamics to effectively engage with the diagnostic instrument. The primary research variables include the development of the DB-CDT diagnostic tool, students' data literacy levels, and their graphical representation abilities. Data collection was executed through teacher interviews, classroom observations, and the administration of the diagnostic test. Tools and materials used include the DB-CDT instrument, expert validation rubrics, and smartphones for the digital implementation of the assessment.

Research Procedure

The research was executed through the five stages of the ADDIE model:

Analysis

This stage involves initial activities to identify problems through interviews and observations. Interviews were conducted with a Physics teacher at a high school in Yogyakarta regarding students' difficulties in learning Physics, specifically the Thermodynamics material. Observations were carried out at the same high school concerning the curriculum used and the trend of students' Physics scores, particularly in the Thermodynamics material. After obtaining this preliminary information, the problem-solving solution was determined through a literature study. Subsequently, an analysis of the feasibility and requirements for the product development was conducted. These requirements included the capacity to address problems in the Physics material, facility support, material suitability (syllabus), and the ability of both teachers and students to operate smartphones.

Design

Based on the analysis results, this phase generated the product blueprint. Key activities included: Initial Q-Matrix Formulation, where every response option, including the distractors, across 90 items was theoretically linked to a specific attribute mastery pattern. Additionally, Item Design was conducted, establishing item specifications (e.g., P-V graph context, type of error represented by the distractors), and determining the Cognitive Diagnostic Model to be utilized, namely the DINA (Deterministic Input, Noisy-

AND Gate) Model, which was selected for its ability to diagnose binary attributes (mastered/unmastered).

Development

The instrument was drafted and subjected to expert validation by three specialists in material, assessment, and data literacy. Following revisions, the instrument underwent empirical testing. Psychometric properties were determined using both Classical Test Theory (CTT) and Item Response Theory (IRT/Rasch Model).

Implementation

The validated and revised DB-CDT product was then pilot tested on a representative sample of high school students in a large-scale testing context. This phase aimed at collecting empirical data regarding student responses to each item and distractor. The test administration was conducted simultaneously, ensuring standardized working conditions for all research subjects, thus making the collected data valid for diagnostic analysis.

Evaluation

The final phase focuses on assessing product quality based on the collected data. Summative Evaluation was performed through classical statistical analysis to calculate reliability (Cronbach's Alpha), discrimination index, and item difficulty level. Subsequently, a diagnostic evaluation was conducted by applying the DINA Model to the student response data and the final Q-Matrix. The results of the DINA analysis were used to (1) estimate the Attribute Mastery Level (AML) of each student, and (2) verify the diagnostic accuracy of the distractors (e.g., whether students who selected distractor A truly failed on attribute A4). The outcomes of this evaluation phase are used to conclude the feasibility and diagnostic effectiveness of the developed DB-CDT. More concisely, the research stages can be seen in the flow diagram shown in Figure 1.

Research Data Analysis

Data analysis was conducted using a multi-method approach to ensure the psychometric integrity of DB-CDT. For the Rasch model analysis, items were considered a good fit if the Item INFIT Mean Square (MNSQ) values ranged between 0.77 and 1.30. This range signifies that the items fit the Rasch model (Subali & Pujiyanti, 2011). The criteria for item retention or rejection after analysis are an INFIT $t \leq 2$. Furthermore, a good item difficulty value ranges between -2 and +2 (Hambleton et al., 1992), while the instrument was deemed reliable if the reliability coefficient reached 0.7 or higher. Distractor effectiveness was evaluated based on a minimum selection rate of 5% by test participants, with a specific focus on ensuring distractors were

selected more frequently by lower-ability students or evenly across ability levels to avoid "misleading" classifications. Final diagnostic accuracy was determined by the alignment between student response patterns and the Q-Matrix via DINA model estimation.

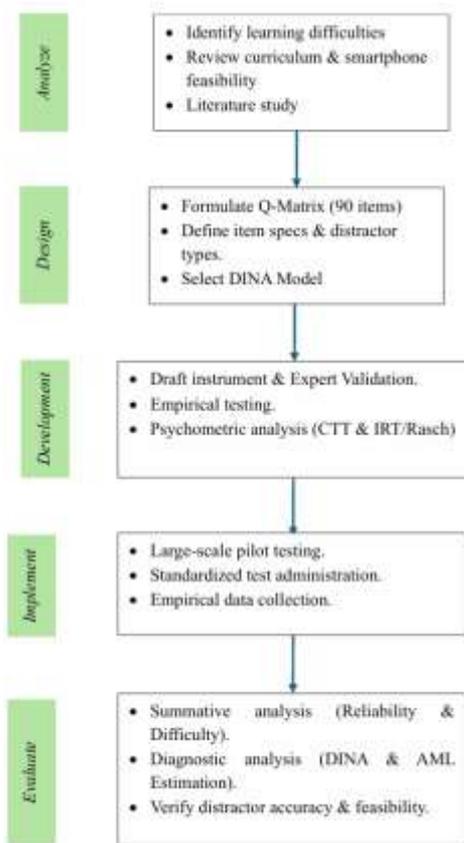


Figure 1. Research flow diagram: ADDIE model

Result and Discussion

Content Validity

The review results consisted of evaluations and input on the media aspect from ten reviewers, comprising two university lecturers, two teachers, and six peer colleagues. The assessment results indicated that the DB-CBT product, in terms of media, could be used effectively. However, there were suggestions from the reviewers, namely that the text size on the website items should be larger and clearer for better visibility by students, and that the scores obtained by each student should be immediately visible upon completion of one diagnostic assessment session. The results of the DB-CBT product assessment by the media and material experts are presented in Table 1.

Overall, all assessment aspects, covering both the media and the material, achieved the maximum V-Aiken value of 1.00, which is consistently interpreted as Excellent. This result indicates that the validated instructional media product possesses the highest level

of validity. Consequently, the media is considered highly feasible and ready for use in the subsequent testing phase, as it has met the quality standards established by experts or validators in terms of design, functionality, and content.

Table 1. Media and Material Evaluation Results of the DB-CBT Product

No Evaluation aspects	Evaluation Items	V-Aiken	Description
Media Review			
1	Product Display	1-21	1.00 Excellent
2	Software Engineering	22-27	1.00 Excellent
Mean		1.00	Excellent
Content Review			
1	Learning	1-6	1.00 Excellent
2	Content	7-30	1.00 Excellent
Mean		1.00	Excellent

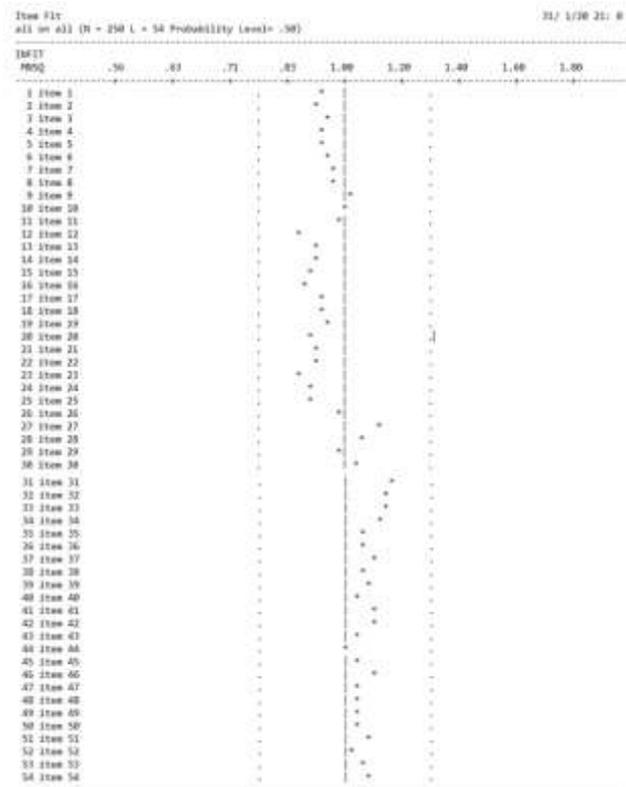


Figure 2. Fit model of data literary items

The empirical analysis of the item quality of the DB-CDT instrument was conducted using the Item Response Theory (IRT) approach via the Quest software. Item quality was assessed based on the goodness of fit value, specifically indicated by the INFIT Mean Square (INFIT MNSQ) statistic. The analysis results indicate that the test items demonstrate an excellent fit within the Fit Model, with a mean INFIT MNSQ value of 1.00 and a standard deviation of 0.08. This mean value is categorized as valid because it falls within the accepted range for the item-fit model, which is between 0.77 and

1.30. Item consistency is also high, evidenced by the item reliability value reaching 0.92, which confirms that the data literacy items in this instrument are reliable. The detailed distribution of the INFIT MNSQ values for each test item is presented in Figure 2, which is part of the analysis results.

The difficulty level of each item is presented by the item difficulty values shown in Figure 3 and Figure 4.

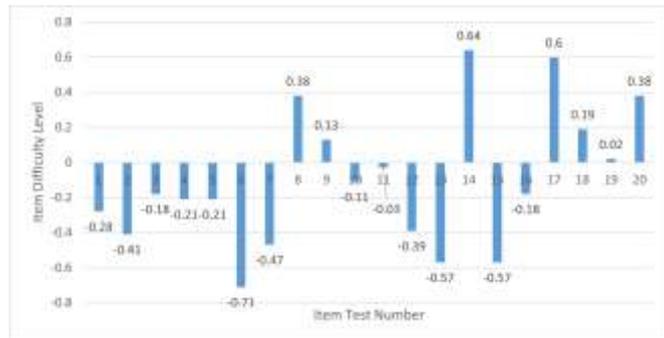


Figure 3. Item difficulty parameter of data literacy questions

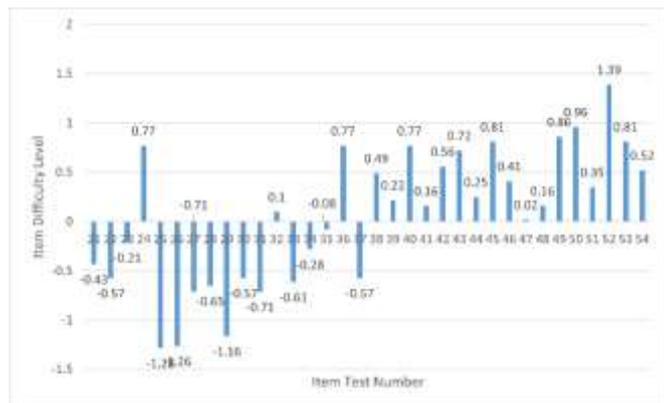


Figure 4. Item difficulty parameter of data literacy questions (continued)

Figures 3 and 4 illustrate that the difficulty level of the data literacy items is categorized as good, as they fall

within the acceptable range of -2 to +2 (Hambleton et al., 1992). The instrument's reliability was assessed using the Item Response Theory (IRT) approach via the total information function and the Standard Error of Measurement (SEM), which provides information concerning the test's reliability. The total information function and the Standard Error of Measurement for the data literacy ability test are presented in Figure 5.

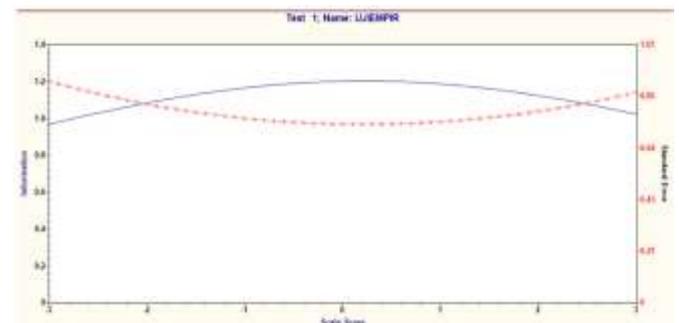


Figure 5. Total information function of data literacy

Figure 5 shows that the test items possess a total information of 1.2. Therefore, the data literacy ability assessment instrument is reliable for measuring the ability of high school students within the interval of -3 to +3, as illustrated in Figure 6.



Figure 6. Histogram of data literacy ability levels

Table 2. Attribute Pattern Q-Matrix

Response Options	Measured Attributes	Required Attribute Pattern (Diagnostic Target)	Distractor Interpretation
Key (D)	A1, A2, A4, A5	(1,1,0,1,0)	Full mastery; drawing the conclusion for the total work done.
Distractor A	A4 Failure	(1,1,0,0,1,0)	Failure in sign convention error (A4): Students fail to apply the correct sign convention.
Distractor B	A2 and A3 Failure	(1,0,0,1,0,0)	Failure in process interpretation (A2 and A3): Students fail to identify the isochoric process (work = 0).
DistractorC (Hypothesis)	A1 Failure	(0,1,1,1,0)	Failure in data reading (A1): Students made an error in reading the variable values (students misread the scale on the PV diagram axes).
Distractor E (Hypothesis)	A5 Failure	(1,1,1,0,1)	Failure in data and concept interpretation (A5): Students are correct in partial calculation but incorrect in summing the total process (ignoring one stage of the cyclic process)

The structure of the DB-CDT is designed not only to measure final outcomes but also to diagnose specific cognitive errors through its distractors, which are linked to the data literacy attributes that students fail to master. The mapping of attribute patterns (Q-Matrix) serves as the diagnostic core. For example, in Item 7, which measures data literacy within the context of Thermodynamics, each option possesses a clear diagnostic meaning, including distractors A and B, for which data are available, and distractors C and E, which are hypothesized.

The diagnostic structure of this instrument is based on the Q-Matrix mapping, where each response option in Item 7 is designed to isolate specific attribute failures: the correct answer (Key D) indicates full mastery with the pattern (1, 1, 0, 1, 1, 0), while Distractor A targets the sign convention error (A4 Failure) and Distractor B targets failure in process interpretation (A2, A3 Failure). Distractor C is hypothesized to diagnose the most basic failure, which is reading explicit data (A1), implying conceptually correct calculation but using incorrect numerical values from the graph. In contrast, Distractor E aims to test the integration of data and concepts (A5), where students fail to integrate all stages of the cycle to calculate the total magnitude. This granular approach to error identification is crucial, as contemporary research in Cognitive Diagnostic Assessment emphasizes that

measuring not just the presence, but the specific type of error is necessary for effective personalized instruction (Chin et al., 2023; Panahi et al., 2022; Yi, 2023).

Furthermore, in the Thermodynamics First Law question, the distractors are designed to differentiate errors. For instance, Distractor B diagnoses failure in applying the volume variable and an error in unit conversion. This highlights the complex interplay between conceptual understanding and procedural fluency, a common difficulty reported in physics education, where students struggle to reconcile macroscopic variables with micro-level unit requirements (Wladis et al., 2020). Meanwhile, Distractor E diagnoses inappropriate equation usage (summing the final and initial volumes) even though the student correctly applied the pressure and change-in-volume variables. This diagnostic system ensures students receive highly specific remediation advice focused on the root of their errors.

An example of the diagnosis for a Thermodynamics First Law item can be demonstrated in Table 3. Another test item, which measures the ability to apply variables from data presented in the context of the Thermodynamics First Law (an isobaric process from 30 m³ to 50 m³ at a pressure of 80 Pa, illustrates how the distractors subtly differentiate error types.

Table 3. Example of the diagnosis for the First Law

Responses Options	Diagnosis Provided	Suggested Remediation
Key (D) (Answer: 1600 J)	Correct response. You are able to apply the pressure variable and the change-in-volume to the equation to determine the work value.	You may proceed to the next topic.
Distractor A (Answer: -1600 J)	Incorrect Response. You were accurate in applying the pressure and change-in-volume variables in the equation to determine the work value. However, you were inaccurate in applying the initial and final volume values within the work equation, specifically that the initial volume value is smaller than the final volume.	Remediation Suggestion: You should revisit the method for applying the initial and final volume variables in the work equation, noting that the initial volume value is smaller than the final volume.
Distractor B (Answer: -1,6 J)	Incorrect response. You were correct in applying the pressure variable in the equation to determine the work value. However, you were inaccurate in applying the initial and final volume values in the work equation, as well as being inaccurate in converting the units used for volume.	You should review how to apply the initial and final volume variables in the work equation; specifically, the initial volume value should be smaller than the final volume, and there is no need to change the unit for the change in volume, as the unit presented in the problem is already appropriate.
Distractor C (Answer: 1,6 J)	Incorrect Answer. You have correctly applied the pressure and change-in-volume variables in the equation to determine the work value. However, you were inaccurate in converting the units used.	You should review the proper method for applying the change-in-volume variable with the appropriate units.
Distractor D (Answer: 6400 J)	Incorrect Answer. You were accurate in applying the pressure and change-in-volume variables in the equation to determine the work value. However, the equation you used was inappropriate, specifically by summing the final and initial volumes.	You should review and relearn the correct application of the equation for determining the work value, which involves multiplying the pressure value by the difference between the final and initial volumes

Table 3 demonstrates the effectiveness of the DB-CDT in isolating the source of student errors. Students who selected Distractor B were diagnosed with a dual error: they failed to apply the volume variable correctly and made an error in unit conversion. Distractor B specifically targets Data Literacy failures related to the use of units and variables. Students who selected Distractor E were diagnosed with using an entirely inappropriate equation (summing the final and initial volumes), even though they may have correctly identified the pressure and volume variables. This distractor directs remediation advice toward multiplying pressure by the change in volume, which is the correct formula. The ability of the DB-CDT to distinguish between these subtly different error types (e.g., calculation vs. conceptual formula usage) underscores the power of Cognitive Diagnostic Assessment (CDA) over traditional scoring methods (Noh et al., 2024). The precise isolation of errors is crucial, as effective instructional design in STEM relies on pinpointing the specific gap in conceptual or procedural knowledge, rather than simply identifying an incorrect result (Bulut et al., 2024). This system

ensures that every student receives a diagnosis and recommendations that differ according to the level and specific type of error they committed – whether related to unit conversion, variable application, or basic formula use – thereby enabling highly focused learning interventions.

The diagnostic assessment was conducted across three sessions, where the instrument measured data literacy ability within the context of Thermodynamics material. While the instrument was developed to measure data literacy, the subsequent reporting of 'digital literacy ability' in Table 4 necessitates careful distinction. Data literacy, focusing on reading, interpreting, and communicating data (Dogan, 2023; Downes, 2023), is a subset often enhanced by digital tools, but the two are not interchangeable. Regardless of the label used in the table, the diagnostic focus remains on the cognitive attributes mapped in the Q-Matrix, which primarily reflect data skills related to scientific problem-solving. The results of the diagnostic assessment, specifically the profile of digital literacy ability, are presented in Table 4.

Table 4. The Diagnostic Assessment of the Digital Literacy Profile

Session	Aspect	Key Answer	Experimental Class (n = 30)	Control Class (n = 30)
1	To collect data	D	86.67% (26 students)	63.3% (19 students)
	To relate multiple data points	A	76.67% (23 students)	80% (24 students)
2	To analyze data	D	96.6% (29 students)	76.6% (23 students)
	To interpret data	A	90% (27 students)	63.3% (19 students)
	To interpret data	A	76.6% (23 students)	70% (21 students)
3	To draw a conclusion	C	76.67% (23 students)	66.6% (20 students)

The results of the diagnostic assessment using the DB-CDT instrument on Thermodynamics material indicate a significant difference in the data literacy ability profile between the experimental class (n=30) and the control class (n=30), spanning six key aspects. Overall, the Experimental Class demonstrated a considerably superior mastery across most of the cognitive attributes measured. The most notable superiority was found in higher-order cognitive abilities such as analyzing data (96.6% vs 76.6%) and interpreting data (90% vs 63.3%), as well as in the basic aspect of collecting data (86.67% vs 63.3%). This large percentage difference (over 20% in these three aspects) strongly supports the effectiveness of the learning intervention applied in the experimental class, which was assumed to focus on enhancing data literacy. This substantial effect size aligns with recent systematic reviews highlighting the positive impact of diagnostic feedback on student performance in STEM fields (Cao et al., 2025). Nevertheless, one diagnostic anomaly was found, where the Control Class was slightly superior in the aspect of relating multiple data points (80% vs 76.67%).

Furthermore, the analysis shows that the primary area of weakness in the control class lies in the inference and synthesis stages, namely in the aspects of interpreting data (lowest at 63.3%) and drawing a conclusion (66.6%). This confirms existing pedagogical challenges, as students often struggle with the complex cognitive load required to infer meaning and synthesize findings from quantitative representations, particularly in conceptually challenging topics like Thermodynamics (Kim et al., 2024). The entirety of these data validates the DB-CDT's capability as a sensitive diagnostic instrument for accurately profiling the specific strengths and weaknesses of high school students in Thermodynamics data literacy.

Conclusion

This Research and Development (R&D) study successfully designed and validated the Distractor-Based Cognitive Diagnostic Test (DB-CDT) for Thermodynamics material, using the ADDIE framework and the DINA Model to specifically assess high school

students' data literacy. The instrument achieved the highest validity from expert evaluation (maximum V-Aiken value of 1.00, interpreted as Excellent) and demonstrated excellent psychometric quality in empirical testing, evidenced by a mean INFIT MNSQ value of 1.00 (indicating a fit with the Rasch model), good item difficulty levels (ranging from -2 to +2), and high item reliability (0.92). Diagnostically, the DB-CDT proved to be sensitive and effective as it is based on a Q-Matrix designed to isolate specific cognitive failures through its distractors (e.g., failure in sign convention or process interpretation), thus enabling focused diagnosis. Analysis of the ability profile showed that the associated learning intervention was effective in significantly enhancing mastery of higher-order cognitive aspects in the experimental class (e.g., analyzing data 96.6%) compared to the control class, thereby validating the DB-CDT's feasibility as an accurate diagnostic tool for profiling students' specific strengths and weaknesses in Thermodynamics data literacy.

Acknowledgments

The author would like to thank the Directorate of Research and Community Service (DRPM), Deputy for Strengthening Research and Development of the Ministry of Research, Technology, and Higher Education/National Research and Innovation Agency for the funding support provided.

Author Contributions

For this article, there are 4 authors. First, S.: conceptualization of the project, the formulation of the main methodology, and overall supervision, as well as the writing and editing of the manuscript draft. Secondly, A.M.B.N.: contributed to data collection and formal analysis, and assisted in the writing of the methods and results sections. Third, R.A.: contributed to the design and validation of the test instrument (DB-CDT), along with a literature review. Finally, P.E.L.: data visualization, the preparation of all illustrations, and the writing of the discussion and conclusion sections. All authors have read and approved the final version of this manuscript and agree to be accountable for all aspects of the work.

Funding

This research was funded by the Directorate of Research and Community Service (DRPM), Deputy for Strengthening Research and Development of the Ministry of Research, Technology, and Higher Education/National Research and Innovation Agency.

Conflicts of Interest

No conflict interest.

References

- Branch, R. M. (2009). *Instructional Design: The ADDIE Approach*. Springer.
- Bulut, O., & Beiting-Parrish, M. (2024). The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges. *Chinese/English Journal of Educational Measurement and Evaluation*, 5(3). <https://doi.org/10.59863/miq17785>
- Cao, X., Lu, H., Wu, Q., & Hsu, Y. (2025). Systematic review and meta-analysis of the impact of STEM education on students learning outcomes. *Frontiers in Psychology*, 16(August), 1–15. <https://doi.org/10.3389/fpsyg.2025.1579474>
- Chin, H., & Chew, C. M. (2023). Validation of cognitive models for subtraction of time involving years and centuries. *International Journal of Assessment Tools in Education*, 10(2), 175–196. <https://doi.org/10.21449/ijate.1160120>
- Davison, M. L., Weiss, D. J., DeWeese, J. N., Ersan, O., Biancarosa, G., & Kennedy, P. C. (2023). A Diagnostic Tree Model for Adaptive Assessment of Complex Cognitive Processes Using Multidimensional Response Options. *Journal of Educational and Behavioral Statistics*, 48(6), 914–941. <https://doi.org/10.3102/10769986231158301>
- 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. <https://doi.org/10.36681/tused.2025.017>
- Dogan, E. (2023). A Meta-Analysis Study on Data Literacy Education For School Administrators and Teachers. *Kuramsal Eğitim Bilim*, 16(1), 199–217. <https://doi.org/10.30831/akukeg.1134207>
- Downes, S. (2023). Three Frameworks for Data Literacy. *20th International Conference on Cognition and Exploratory Learning in Digital Age*, 107–115. https://doi.org/10.33965/celda2023_2023061014
- Guglietti, M. V. (2023). Redefining visual literacy in an era of visual overload: The use of reflective visual journals to expand students' visual thinking. *Journal of University Teaching and Learning Practice*, 20(4). <https://doi.org/10.53761/1.20.4.03>
- Hambleton, R. K., Swaminathan, H., & Rogers, D. J. (1992). Fundamentals of item response theory. *Choice Reviews Online*, 29(07), 29-4185-29-4185. <https://doi.org/10.5860/CHOICE.29-4185>
- Kim, M., Duncan, C., Yip, S., Sankey, D., Kim, M., Duncan, C., Yip, S., & Dec, D. S. (2024). Beyond the theoretical and pedagogical constraints of cognitive load theory, and towards a new cognitive philosophy in education. *Educational Philosophy and Theory*, 1(1), 1–12. <https://doi.org/10.1080/00131857.2024.2441389>
- Laliyo, L. A. R., La Kilo, A., Paputungan, M., Kunusa, W. R., Dama, L., & Panigoro, C. (2022). Rasch Modelling To Evaluate Reasoning Difficulties, Changes of Responses, and Item Misconception

- Pattern of Hydrolysis. *Journal of Baltic Science Education*, 21(5), 817–835. <https://doi.org/10.33225/jbse/22.21.817>
- Laliyo, L. A. R., Tangio, J. S., Sumintono, B., Jahja, M., & Panigoro, C. (2020). Analytic approach of response pattern of diagnostic test items in evaluating students' conceptual understanding of characteristics of particle of matter. *Journal of Baltic Science Education*, 19(5), 824–841. <https://doi.org/10.33225/jbse/20.19.824>
- Lee, J., Alonzo, D., Beswick, K., Abril, J. M. V., Chew, A. W., & Oo, C. Z. (2024). Dimensions of teachers' data literacy: A systematic review of literature from 1990 to 2021. In *Educational Assessment, Evaluation and Accountability* (Vol. 36, Issue 2, pp. 145–200). Springer Netherlands. <https://doi.org/10.1007/s11092-024-09435-8>
- Li, Y. (2023). Impact of perceived recognition by physics instructors on women 's self-efficacy and interest. *Physical Review Physics Education Research*, 19(2), 20125. <https://doi.org/10.1103/PhysRevPhysEducRes.19.020125>
- Lian, L. H., & Yew, W. (2023). Determining the psychometric properties of middle school statistical thinking testlet-based assessment tool. *International Journal of Assessment Tools in Education*, 10(4), 672–689. <https://doi.org/10.21449/ijate.1255859>
- Mandinach, E. B., & Schildkamp, K. (2021). Misconceptions about data-based decision making in education: An exploration of the literature. *Studies in Educational Evaluation*, 69(September), 100842. <https://doi.org/10.1016/j.stueduc.2020.100842>
- Mulyani, S., Haniza, N., Ramadhani, D. G., & Mahardiani, L. (2021). Rash Model Approach for Analysis of Misconception on Chemistry Learning with Distractor Analysis. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 6(1), 98. <https://doi.org/10.20961/jkpk.v6i1.50517>
- Munandar, H., Doyan, A., Susilawati, S., Hakim, S., Mulyadi, L., & Hamidi, H. (2024). Increasing Motivation to Study Physics Using PhET Media on Mechanical Energy Material. *MANDALIKA : Journal of Social Science*, 2(1), 1–5. <https://doi.org/10.56566/mandalika.v2i1.70>
- Natalis, V., & Leyh, B. (2025). How Difficult is it to teach Thermodynamics in Tertiary Education? Teachers' Opinions and Predominance of the Disorder Metaphor. *International Journal of Research in Education and Science*, 11(2), 190–215. <https://doi.org/10.46328/ijres.3570>
- Net, W. W. W. P. (2023). Exploration of Diagnostic Testing Instruments: Validity, Reliability, and Item Characteristics. *Pegem Journal of Education and Instruction*, 13(3), 386–394. <https://doi.org/10.47750/pegegog.13.03.39>
- Noh, M. F. M., Mohd Matore, M. E., Sulaiman, N. A., Azeman, M. T., Ishak, H., Othman, N., Mohd Rosli, N., & Sabtu, S. H. (2024). Cognitive Diagnostic Assessment in Educational Testing: A Score Strategy-Based Evaluation. *E-Bangi Journal of Social Science and Humanities*, 20(4). <https://doi.org/10.17576/ebangi.2024.2104.21>
- OECD. (2023). *PISA 2022 Results: The State of Learning and Equity in Education: Vol. I*. Paris: OECD Publishing.
- Panahi, A., & Mohebbi, H. (2022). Cognitive Diagnostic Assessment of IELTS Listening: Providing Feedback from its Internal Structure. *Language Teaching Research Quarterly*, 29, 147–160. <https://doi.org/10.32038/ltrq.2022.29.10>
- Prasanth, P., Reshma, P., & Udayanandan, K. M. (2023). Thermodynamics in Microcanonical Ensemble. *European Journal of Physics Education*, 14(1), 14–22. Retrieved from <https://files.eric.ed.gov/fulltext/EJ1413982.pdf>
- Pusporini, W., Rosana, D., & Setyawarno, D. (2026). Innovative Online Diagnostic Assessment for Evaluating Science Learning Readiness in the Merdeka Curriculum Framework. *Journal of Education and Teaching (JET)*, 7(1), 1–19. <https://doi.org/10.51454/jet.v7i1.546>
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic Measurement (Theory, Methods, and Applications)*. New York: The Guilford Press.
- Scaria, N., John, S., Kennedy, J., Agarwal, K., Seth, D., & Subramani, D. (2025). Dissecting Physics Reasoning in Small Language Models: A Multi-Dimensional Analysis from an Educational Perspective. *ArXiv*. <https://doi.org/2505.20707v2> [cs.CL] null
- Soeharto, S. (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>
- Suryadi, Mahardika, I. K., Supeno, & Sudarti. (2021). Data literacy of high school students on physics learning. *Journal of Physics: Conference Series*, 1839(1), 012025. <https://doi.org/10.1088/1742-6596/1839/1/012025>
- Wladis, C., Verkuilen, J., McCluskey, S., Offenholley, K., Licwinko, S., Lee, J. K., Wladis, C., Verkuilen, J., McCluskey, S., Offenholley, K., Dawes, D., Wladis, C., Verkuilen, J., McCluskey, S., Offenholley, K., & Dawes, D. (2020). Relationships between procedural fluency and conceptual understanding in algebra for postsecondary students To cite this

version: HAL Id: hal-02416499 Relationships between procedural fluency and conceptual understanding in algebra for postsecondary stu. In *Eleventh Congress of the European Society for Research in Mathematics Education*. Retrieved from <https://hal.science/hal-02416499/>

Yi, Y. S. (2023). Another Reason for CDA in Language Assessment: A Critical Synthesis from a Perspective of Validating Test Constructs. *Language Teaching Research Quarterly*, 37, 248–265. <https://doi.org/10.32038/ltrq.2023.37.14>