Jurnal Pijar MIPA
https://jurnalfkip.unram.ac.id/index.php/JPM
ISSN 1907-1744 (Print)
ISSN 2460-1500 (Online)

Development of a Critical Thinking Skills Test Instrument Based on Rasch Modelling for Temperature and Heat

Elmerita Dwi Agustin*, Hera Novia, Raden Giovanni Ariantara

Department of Physics Education, Faculty of Mathematics and Natural Sciences, Indonesia University of Education, Bandung, Indonesia
*e-mail: immerikwon@gmail.com

Received: May 9, 2026. Accepted: May 30, 2026. Published: July 9, 2026

Abstract: Critical thinking skills are essential competencies in 21st-century physics learning. However, the availability of assessment instruments that accurately measure these skills remains limited, particularly for physics topics such as temperature and heat. Previous studies have rarely integrated Rasch modelling into the development of critical-thinking assessment instruments, particularly at the secondary school level. Therefore, this study aimed to develop and analyze the validity, reliability, and item fit of a critical thinking skills test instrument on temperature and heat topics using the Rasch model. This study employed a Research and Development (R&D) approach using the ADDIE model, comprising Analysis, Design, Development, Implementation, and Evaluation stages. The instrument consisted of 20 multiple-choice items constructed from critical-thinking indicators. Content validity was analyzed using Aiken's V, while empirical analysis was conducted using the Rasch model through Ministeps Software. The analysis included item reliability, person reliability, item fit, difficulty level, and unidimensionality. The results showed that the instrument achieved an average Aiken's V value of 0.8068, indicating that the instrument met the content validity criteria. Rasch analysis revealed that the item reliability was 0.86, while the person reliability was 0.35. In addition, 19 items were categorized as fit, whereas one item was identified as a misfit and required revision. The instrument also met the unidimensionality requirement with a raw variance explained value of 22.7%. In conclusion, the developed instrument meets the criteria for validity, reliability, and item fit and is suitable for measuring students' critical thinking skills on topics related to temperature and heat. The findings indicate that Rasch modelling can support the development of more objective and comprehensive assessment instruments for physics learning and classroom assessment practices.

Keywords: Assessment; Critical Thinking Skills; Rasch Model; Temperature and Heat; Test Instrument.

Introduction

Critical thinking skills are one of the essential competencies in 21st-century learning that students must possess. These skills play a crucial role in enabling students to understand concepts deeply, analyze information, evaluate arguments, and make rational, evidence-based decisions [1] [2]. In the context of physics education, critical thinking skills are particularly important because students are required to comprehend abstract concepts and solve problems systematically and logically [3] [4]. In physics learning, critical thinking skills also help students connect scientific concepts to real-life phenomena and solve problems systematically [5].

Recent studies indicate that students' critical thinking skills remain relatively underdeveloped [6]. This condition is attributed to teacher-centered instruction and the lack of evaluation tools capable of measuring higher-order thinking skills [7]. Recent research also indicates that integrating technology and digital learning media can enhance students' critical thinking skills; however, implementation remains suboptimal in school settings [8] [9].

A similar situation is also found in Indonesia. Several studies report that students' critical thinking skills in physics education still need improvement through the implementation of innovative learning strategies and appropriate assessment instruments [10]. More recent studies also emphasize that the use of Higher Order Thinking Skills (HOTS)-based instruments and the problem-based learning (PBL) approaches can significantly improve students' critical thinking skills [11] [12]. Therefore, assessment instruments that can objectively and comprehensively measure critical thinking skills are needed to provide more accurate information about students' abilities [13], [14].

However, the development of instruments to measure critical thinking skills continues to face various challenges [15]. One of the main problems is the use of classical test theory approaches, which have limitations in deeply analyzing item quality, such as reliability, difficulty level, and item fit [16] [17]. This approach often fails to provide comprehensive and accurate information about the instrument's overall characteristics [18].

As an alternative, the Rasch model offers a modern measurement approach that enables more comprehensive analysis of instrument characteristics, including reliability, item difficulty, and item fit [19]. This model allows researchers to evaluate reliability, difficulty level, and item fit simultaneously, thereby producing more valid and accurate instruments [20]. Additionally, the Rasch model enables unidimensionality analysis to ensure that the instrument truly measures a single main construct [17].

How to Cite:
E. D. Agustin, H. Novia, and R. G. Ariantara, "Development of a Critical Thinking Skills Test Instrument Based on Rasch Modelling for Temperature and Heat", *J. Pijar.MIPA*, vol. 21, no. 4, pp. 678–684, Jul. 2026. https://doi.org/10.29303/jpm.v21i4.12069

Therefore, the Rasch model has been widely used in educational measurement research because it provides more detailed measurement information than classical test theory approaches [21], [22].

Although various studies have demonstrated the advantages of the Rasch model, its application in developing critical thinking skill instruments for physics education, particulary regarding temperature and heat, remains limited [23]. Moreover, studies that analyze validity, reliability, and item fit of instruments using the Rasch model in depth at the secondary school level are still scarce [24]. This condition indicates the need for assessment instruments that can measure critical thinking skills more objectively and in ways appropriate to the characteristics of physics learning [25], [26].

Based on the above considerations, there is a gap between the need for critical thinking assessment instruments that meet the criteria of validity, reliability, and item fit and the availability of instruments that have been comprehensively tested [27]. Therefore, this study aims to develop and analyze the validity, reliability, and item fit of a test instrument for measuring students' critical thinking skills on temperature and heat topics using the Rasch model.

Research Methods

This study employs a Research and Development (R&D) approach to develop a valid and reliable instrument for assessing critical thinking skills on the topic of temperature and heat [28]. The R&D approach was chosen because this study aims not only to measure but also to produce an educational instrument through a systematic process of design, development, validation, and evaluation [16]. The research stages include needs analysis, instrument design, item development, expert validation, revision, pilot testing, data analysis, and instrument evaluation and finalization [29].

At the Analysis stage, a needs analysis was conducted through literature reviews and classroom observations to identify problems related to measuring students' critical thinking skills in physics learning [15]. In the design stage, the instrument blueprint was prepared, critical thinking indicators were determined, and item specifications were developed [30], [31]. The Development stage included developing the instrument items, conducting expert validation, and revising the instrument based on the validators' suggestions [32]. The Implementation stage involved conducting the field testing with students, while the Evaluation stage involved analyzing the collected data using the Rasch model [33].

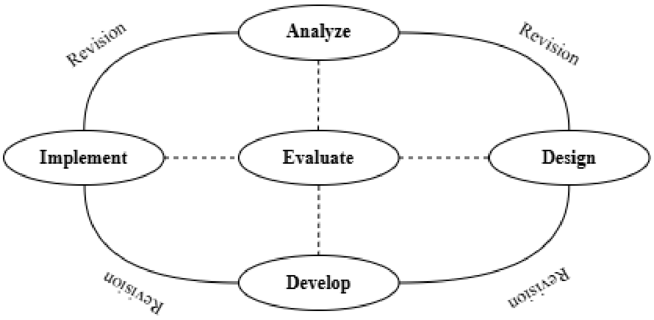


Figure 1. ADDIE-based instrument development procedure

The research subjects were 11th-grade high school students who had studied temperature and heat. Subject selection was conducted using a purposive approach, considering the alignment of the material studied with the developed instrument [34]. The research instrument consists of 20 multiple-choice items designed based on critical thinking skill indicators according to [35], which include analyzing arguments, identifying assumptions, evaluating information, drawing conclusions, and making decisions.

Data collection was conducted in several stages. In the initial stage, a needs analysis was performed to identify issues in measuring students' critical thinking skills through literature reviews and observations [15]. Subsequently, expert validation was conducted using an instrument validation sheet covering aspects of content, construction, and language. Each item was assessed using a Likert scale ranging from 1-to-5, from "Not Applicable" to "Very Applicable" [36]. This validation process aimed to ensure that the developed instrument was appropriate for measuring indicators of critical thinking skills and suitable for research use [37].

Table 1. Likert Scale

Criteria	Description
1	Very Appropriate
2	Suitable
3	Fairly Appropriate
4	Not Very Appropriate
5	Not Appropriate

The expert validation results were analyzed using Aiken's V index to determine the instrument's content validity. The instrument was validated by five validators, consisting of three physics education lecturers and two physics teachers. Revisions were subsequently made based on validators' suggestions before the instrument was tested with students [38], [39]. The revisions included improving item wording, clarifying contextual information, refining answer options, and adjusting items to better align with critical-thinking indicators. In addition, validators suggested adding image identifiers and shortening several item narratives to reduce students' misconceptions and improve readability [40], [41]. One example of item revision based on validator feedback is presented in Figures 2 and 3.

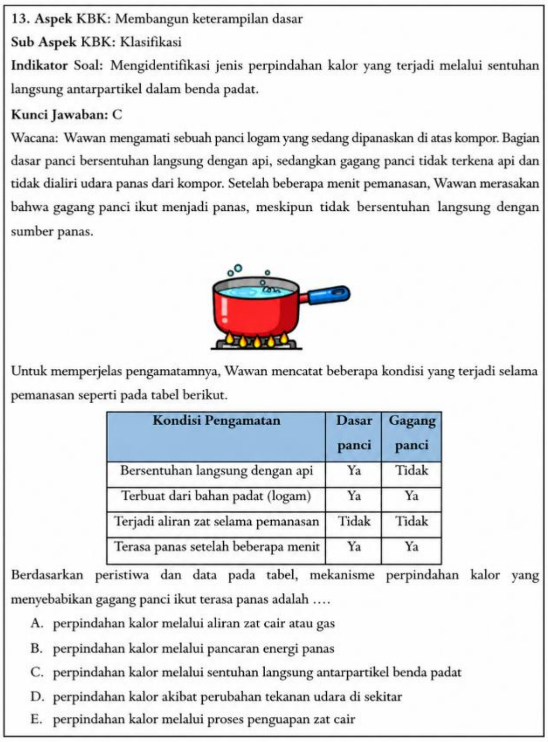
Table 2. Aiken's V Index

Criteria	Description
$V \geq 0.80$	Valid
$V < 0.80$	Invalid

The validity formula proposed by Aiken is as follows:

$$V = \frac{\sum s}{n(c-1)}$$

Explanation:
V: Item validity index
$s = r - l_0$: Score given by the rater;
l_0: Lowest score
c : Number of rating categories
n : Number of validators

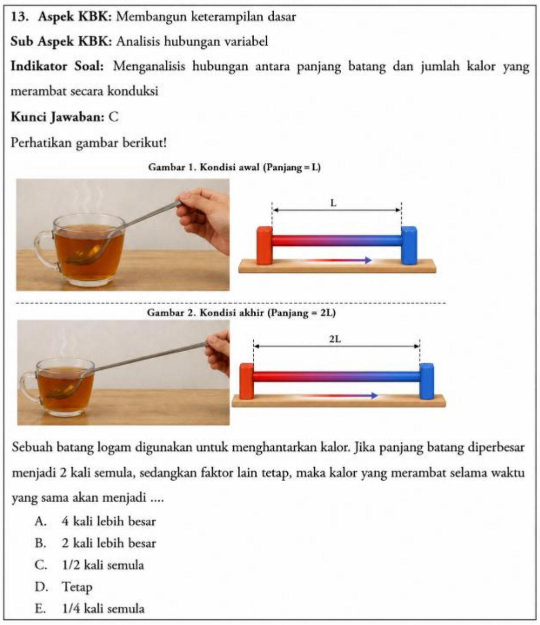

13. Aspek KBK: Membangun keterampilan dasar
Sub Aspek KBK: Klasifikasi
Indikator Soal: Mengidentifikasi jenis perpindahan kalor yang terjadi melalui sentuhan langsung antarpartikel dalam benda padat.
Kunci Jawaban: C
Wacana: Wawan mengamati sebuah panci logam yang sedang dipanaskan di atas kompor. Bagian dasar panci bersentuhan langsung dengan api, sedangkan gagang panci tidak terkena api dan tidak dialiri udara panas dari kompor. Setelah beberapa menit pemanasan, Wawan merasakan bahwa gagang panci ikut menjadi panas, meskipun tidak bersentuhan langsung dengan sumber panas.

Untuk memperjelas pengamatannya, Wawan mencatat beberapa kondisi yang terjadi selama pemanasan seperti pada tabel berikut.

Kondisi Pengamatan	Dasar panci	Gagang panci
Bersentuhan langsung dengan api	Ya	Tidak
Terbuat dari bahan padat (logam)	Ya	Ya
Terjadi aliran zat selama pemanasan	Tidak	Tidak
Terasa panas setelah beberapa menit	Ya	Ya

Berdasarkan peristiwa dan data pada tabel, mekanisme perpindahan kalor yang menyebabikan gagang panci ikut terasa panas adalah ....
A. perpindahan kalor melalui aliran zat cair atau gas
B. perpindahan kalor melalui pancaran energi panas
C. perpindahan kalor melalui sentuhan langsung antarpartikel benda padat
D. perpindahan kalor akibat perubahan tekanan udara di sekitar
E. perpindahan kalor melalui proses penguapan zat cair

Figure 2. Example of a test item before revision


13. Aspek KBK: Membangun keterampilan dasar
Sub Aspek KBK: Analisis hubungan variabel
Indikator Soal: Menganalisis hubungan antara panjang batang dan jumlah kalor yang merambat secara konduksi
Kunci Jawaban: C
Perhatikan gambar berikut!

Sebuah batang logam digunakan untuk menghantarkan kalor. Jika panjang batang diperbesar menjadi 2 kali semula, sedangkan faktor lain tetap, maka kalor yang merambat selama waktu yang sama akan menjadi ....
A. 4 kali lebih besar
B. 2 kali lebih besar
C. 1/2 kali semula
D. Tetap
E. 1/4 kali semula

Figure 3. Example of an item after revision

Following the revisions, the developed items became clearer, more systematic, and more closely aligned with the critical thinking indicators being assessed [42], [43], [44], [45]. The revision process also contributed to minimizing potential misconceptions and enhancing the readability of the instrument prior to the field testing stage.[46].

The instrument trial was conducted in one testing session under the supervision of the physics teacher to obtain students' response data [47], [48], [49]. The collected data were analyzed using the Rasch model through Winsteps software because this model provides more comprehensive information regarding instrument characteristics compared to classical test theory [20]. The analysis included person and item reliability, item fit based on INFIT and OUTFIT Mean Square (MNSQ) values, difficulty levels on the logit scale, and unidimensionality through Principal Component Analysis (PCA) of the residuals [50]. The interpretation of the analysis results was based on the Rasch model criteria, namely:

Table 3. Item Fit Criteria Based on the Rasch Model

Criteria	Value
Outfit Mean Square (MNSQ) Value	0.5 < MNSQ < 1.5
Z-Standard Outfit Value (ZSTD)	-2.0 < ZSTD < +2.0
Point Measure Correlation Value (PtMeaCorr)	0.4 < PtMeaCorr < 0.85

Table 4. Reliability Criteria for the Rasch Model

Parameter	Value	Category
Person Reliability	> 0.80	Very Good
	0.60 – 0.80	Good
	0.40 – 0.60	Fair
	< 0.40	Weak
Reliability Item	>0.90	Very Good
	0.80 – 0.90	Good
	0.70 – 0.80	Fair
	< 0.70	Low
Cronbach's Alpha	> 0.80	Very Good
	0.60 – 0.80	Good
	0.40 – 0.60	Fair
	< 0.40	Low

Results and Discussion

The results of this study present the validity, reliability, item fit, difficulty level, and unidimensionality of the critical thinking skills test instrument, developed through expert validation and Rasch model analysis [51], [52]. The instrument consisted of 20 multiple-choice items on temperature and heat and was developed based on critical-thinking indicators proposed by [53].

In the initial stage, the instrument was validated by five expert validators, comprising three physics education lecturers and two physics teachers. Validation was conducted for the aspects of content, construction, and language using a 1-5 Likert scale [54]. The results of Aiken's V index calculation for each item and their averages are presented in Table 5.

Table 5. Results of Content Validity for the Critical Thinking Skills Test Instrument (n = 5 validators)

Aspect	Average Aiken's V	Category
Content	0.8121	Valid
Construction	0.8046	Valid
Linguistics	0.8038	Valid
Overall Average	0.8068	Valid

Based on Table 5, the average Aiken's V index for all aspects is 0.8068, which falls within the valid category (V ≥ 0.80). These results indicate that the developed instrument met the content validity criteria and was suitable for field testing after revisions based on validators' suggestions. The highest validation score was obtained for the content aspect, indicating that the developed items were relevant to the critical-thinking indicators and to the concepts of temperature and heat. The Aiken's V value for each item is presented in Table 6.

Table 6. Aiken's V Values for Each Item

Item	Aiken's V	Item	Aiken's V	Item	Aiken's V
Item 1	0.800	Item 8	0.812	Item 15	0.800
Item 2	0.812	Item 9	0.804	Item 16	0.800
Item 3	0.808	Item 10	0.804	Item 17	0.817
Item 4	0.804	Item 11	0.800	Item 18	0.796
Item 5	0.804	Item 12	0.817	Item 19	0.800
Item 6	0.800	Item 13	0.808	Item 20	0.800
Item 7	0.804	Item 14	0.812		

The results in Table 6 show that all items achieved Aiken's V value above 0.80, indicating that each item met the validity criteria. These findings suggest that the developed items were considered appropriate in terms of content, construction, and language for measuring students' critical thinking skills [27]. Revisions based on validators' suggestions were made before the field testing stage to improve the instrument's clarity and readability. Examples of item revisions are presented in Figure 2 and Figure 3. After the revision process, the instrument was tested with 11th-grade students who had studied the topics of temperature and heat [55]. The collected response data were analyzed using the Rasch model through Winsteps software. The reliability analysis results are presented in Table 7.

Table 7. Reliability Analysis Results Based on the Rasch Model

Analysis Aspect	Value	Category
Person Reliability	0.35	Weak
Item Reliability	0.86	Good
Cronbach's Alpha	0.51	Moderate

Based on Table 7, the item reliability value of 0.86 indicates that the developed items showed good consistency in measuring critical thinking skills. This finding is consistent with previous studies reporting that item reliability above 0.80 indicates good consistency of instrument items in measuring the intended construct [17], [56]. This result suggests that the items functioned consistently and adequately differentiated students' abilities. Meanwhile, the person's reliability value of 0.35 was categorized as weak, indicating variations in students' response patterns and differences in students' critical thinking abilities. This result may have been influenced by differences in students' understanding of temperature and heat concepts, as well as by students' limited experience answering critical-thinking-based questions.

The Cronbach's Alpha value of 0.51 indicates a moderate interaction between persons and items. These findings are consistent with previous Rasch-based studies indicating that low person reliability can result from differences in students' abilities, inconsistent response patterns, and limited familiarity with higher-order thinking questions. The item fit analysis was conducted to determine the suitability of each item with the Rasch model criteria. The results of the item fit analysis are presented in Table 8.

Table 8. Item Fit Analysis Results

Category	Number of Items	Description
Fit	19	Matches
Not Fit	1 (Item 13)	Needs Revision

The analysis showed that 19 items met the Rasch model fit criteria, indicating that the items functioned consistently in measuring students' critical thinking skills. However, one item was categorized as a misfit and required revision. The presence of a misfit item indicates inconsistency between the item and students' response patterns. Therefore, revision was necessary to improve the alignment between the item and the measured construct.

These findings indicate that most of the developed items were appropriate for measuring critical thinking skills and aligned with the Rasch measurement model. Similar findings have also been reported in previous Rasch-based instrument development studies, in which a small number of misfit items commonly appear during testing and require refinement before wider implementation. The difficulty-level analysis was conducted using logit values from the Rasch analysis. The distribution of item difficulty levels is presented in Table 10.

Table 9. Distribution of Item Difficulty Levels

Difficulty Level	Number of Items
Very Easy (Logit ≤ -1.0)	Items 17, 19
Easy (-1.0 < Logit ≤ -0.5)	Items 2, 5, 8, 9, 12, 14, 15, 18, 20
Moderate (-0.5 < Logit ≤ 0.5)	Items 1, 3, 6, 7, 10, 16
Difficult (0.5 < Logit ≤ 1.0)	-
Very Difficult (Logit > 1.0)	Items 4, 11

Based on Table 9, the results showed that the developed instrument consisted of items with varying difficulty levels. Two items were categorized as very easy, nine items as easy, six items as moderate, and two items as very difficult, while no items were classified as difficult. The variation in item difficulty indicates that the instrument measured students' critical thinking skills across different ability levels [18]. The distribution of item difficulty levels is illustrated in Figure

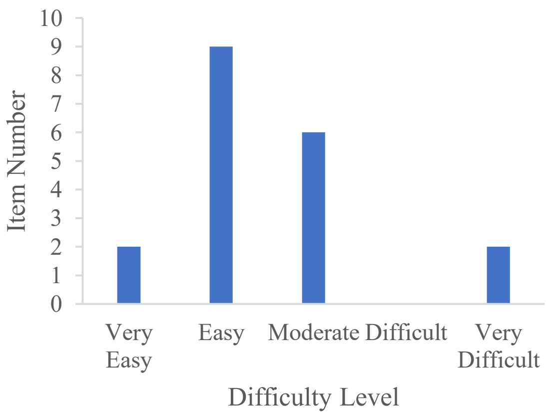


Figure 4. Distribution of item difficulty levels

The presence of items with varying levels of difficulty is important in critical thinking assessment because it enables the instrument to identify students with different levels of understanding and reasoning abilities. In addition, balanced item difficulty contributes to more comprehensive measurement results. The unidimensionality analysis was conducted using Principal Component Analysis (PCA) of residuals to determine whether the instrument measured a single construct consistently. The results are presented in Table 10.

Table 10. Unidimensionality Analysis Results

Aspect	Value
Raw Variance Explained by Measures	22.7%

The PCA results showed that the raw variance explained by measures was 22.7%, indicating that the instrument met the unidimensionality requirement. This result suggests that the instrument consistently measured a single construct, namely, students' critical thinking skills on temperature and heat topics. The obtained value also indicates that the instrument measured a single dominant construct, which is important in Rasch-based measurement analysis.

Overall, the findings indicate that the developed instrument met the criteria of validity, reliability, item fit, and unidimensionality based on Rasch model analysis [56]. Therefore, the instrument can be used to measure students' critical thinking skills in physics, particularly in the topics of temperature and heat. In addition, the use of the Rasch model contributed to a more objective and comprehensive analysis in evaluating the developed instrument.

Conclusion

This study successfully developed a critical-thinking skills test instrument on the topics of temperature and heat using the Rasch model. The instrument was developed using the ADDIE model, including Analysis, Design, Development, Implementation, and Evaluation. The expert validation results showed that the instrument achieved an average Aiken's V of 0.8068, indicating it met the content validity criteria. In addition, the Rasch analysis results showed that the instrument had good item reliability with a value of 0.86, while the person reliability value was categorized as weak at 0.35. The item fit analysis indicated that 19 items met the Rasch model fit criteria, whereas one item required revision. Furthermore, the instrument met the unidimensionality requirement with a raw variance explained value of 22.7%. The developed instrument also consisted of items with varying levels of difficulty, including very easy, easy, moderate, and very difficult categories, indicating that the instrument was able to measure students' critical thinking skills across different ability levels. Overall, the findings indicate that the developed instrument can be used to measure students' critical thinking skills in physics, particularly in the topics of temperature and heat. In addition, the use of the Rasch model contributed to a more objective and comprehensive analysis in evaluating the developed instrument and can support assessment practices in physics education.

Author's Contribution

E.D. Agustin: conceptualization, methodology, instrument development, data collection, data analysis, and writing original draft. H. Novia: supervision, validation, and manuscript review. R.G. Ariantara: supervision, validation, and manuscript review.

Acknowledgements

The authors would like to express their gratitude to the supervisors for their guidance, to the validators for their valuable feedback, and to all students who participated in this study.

References

[1] D. T. Tiruneh, M. De Cock, and J. Elen, "Designing Learning Environments for Critical Thinking: Examining Effective Instructional Approaches," *Int. J. Sci. Math. Educ.*, vol. 16, no. 6, pp. 1065–1089, 2018, doi: 10.1007/s10763-017-9829-z.

[2] R. H. Ennis, "The Nature of Critical Thinking: An Outline of Critical Thinking Dispositions and Abilities i," May 2011.

[3] K. J. Arni, "Systematic Review of Research Trends on Critical Thinking Skills in Physics Learning," *International Journal of Science Education and Science*, vol. 2, no. 1, pp. 43–52, Mar. 2025, doi: 10.56566/ijses.v2i1.262.

[4] G. Lecornu, "Critical Thinking across the Curriculum: A Guide for Teachers.," 2020.

[5] I. Kadek Suarjana, "How Critical Thinking Skills Can Speed Up Solving Physics Problems for Senior High School. SEARCH: Science Education Research," Oct. 2023. doi: https://doi.org/10.47945/search.v2i1.1313.

[6] S. L. E. W. Fajari and Chumdari, "Critical Thinking Skills And Their Impacts On Elementary School Students," *Malaysian Journal of Learning and Instruction*, vol. 18, no. 2, pp. 161–187, Jul. 2021, doi: 10.32890/mjli2021.18.2.6.

[7] K. Arafah, B. D. Amin, S. S. Sari, and A. Hakim, "The Development of Higher Order-Thinking Skills (HOTS) Instrument Assessment in Physics Study," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, May 2021. doi: 10.1088/1742-6596/1899/1/012140.

[8] D. Rahmawati and B. Santoso, "Evaluasi Penggunaan Game Sebagai Media Pembelajaran Fisika: Dampak Terhadap Pemahaman Konsep Siswa SMA," *Jurnal Penelitian Pendidikan*, vol. 24, no. 1, pp. 59–68, 2020.

[9] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Comput. Educ.*, vol. 147, Apr. 2020, doi: 10.1016/j.compedu.2019.103778.

[10] D. A. Kurniawan, A. Astalini, and D. K. Sari, "An evaluation analysis of students' attitude towards physics learning at senior high school," *Jurnal Penelitian dan Evaluasi Pendidikan*, vol. 23, no. 1, pp. 26–35, Jun. 2019, doi: 10.21831/pep.v23i1.20821.

[11] D. A. Kusumaningtyas, M. Manyunu, E. Kurniasari, A. N. Awalin, R. Rahmaniati, and A. Febriyanti, “Enhancing Learning Outcomes: A Study on the Development of Higher Order Thinking Skills based Evaluation Instruments for Work and Energy in High School Physics,” *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, pp. 14–31, Jan. 2024, doi: 10.23917/ijolae.v6i1.23125.

[12] S. M. Brookhart, “How to Assess Higher-Order Thinking Skills in Your Classroom,” 2010. [Online]. Available: www.ascd.org/memberbooks

[13] H. A. Rahman, B. S. Bachri, and I. Y. Maureen, “Instrumen Tes untuk Mengukur Keterampilan Berpikir Kritis pada Materi Statistika,” *Jurnal Cendekia : Jurnal Pendidikan Matematika*, vol. 8, no. 2, pp. 1472–1480, Jul. 2024, doi: 10.31004/cendekia.v8i2.3343.

[14] A. Khasanah, “pengembangan instrumen penilaian otentik pembelajaran fisika untuk mengukur keterampilan proses sains dan berpikir kritis”, [Online]. Available: https://journal.student.uny.ac.id/ojs/index.php/pfisika/index

[15] Y. B. Bhakti, R. Arthur, and Y. Supriyati, “Development of an assessment instrument for critical thinking skills in Physics: A systematic literature review,” in *Journal of Physics: Conference Series*, Institute of Physics, 2023. doi: 10.1088/1742-6596/2596/1/012067.

[16] H. Retnawati, *Analisis Kuantitatif Instrumen Penelitian*, Pertama. Yogyakarta, 2016. [Online]. Available: www.nuhamedika.gu.ma

[17] T. G. . Bond and C. M. . Fox, *Applying the Rasch model : fundamental measurement in the human sciences*. Routledge, Taylor and Francis Group, 2015.

[18] C. O. Solomon and O. O. Romy, “application of classical test theory as linear modeling to test item development and analysis,” *International Research Journal of Modernization in Engineering Technology and Science*, no. 10, Oct. 2023, doi: 10.56726/irjmets45379.

[19] B. Trevor G and F. Christine M, “Applying the Rasch Model: Fundamental Measurement in the Human Sciences Second Edition,” 2007.

[20] B. Sumintono and W. Widhiarso, “Aplikasi Model Rasch Untuk Penelitian Ilmu-Ilmu Sosial,” 2014.

[21] D. Hope *et al.*, “Exploring the use of Rasch modelling in ‘common content’ items for multi-site and multi-year assessment,” *Advances in Health Sciences Education*, vol. 30, no. 2, pp. 427–438, Apr. 2025, doi: 10.1007/s10459-024-10354-y.

[22] M. Polat, N. S. Turhan, and Ç. Toraman, “Comparison of Classical Test Theory vs. Multi-Facet Rasch Theory,” *Pegem Egitim ve Ogretim Dergisi*, vol. 12, no. 2, pp. 213–225, 2022, doi: 10.47750/pegegog.12.02.21.

[23] A. Alyamuari, W. Liliawati, and D. Saepuzaman, “item analysis of the critical thinking test instrument on temperature and heat using rasch modelling,” *Jurnal Eduscience (JES)*, Aug. 2024, doi: https://doi.org/10.36987/jes.v11i2.5676.

[24] E. Avinç and F. Doğan, “Digital literacy scale: Validity and reliability study with the rasch model,” *Educ. Inf. Technol. (Dordr).*, vol. 29, no. 17, pp. 22895–22941, Dec. 2024, doi: 10.1007/s10639-024-12662-7.

[25] Khaeruddin and B. D. Amin, “Development and Validation of Critical Thinking Skills Instruments of Physics Subjects,” in *Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019*, European Alliance for Innovation, 2020. doi: 10.4108/eai.12-10-2019.2296474.

[26] B. Sartono, W. Sunarno, B. A. Prayitno, and N. Y. Indriyanti, “Development and Validation of A Critical Thinking Assessment on Temperature and Heat for Secondary Physics Education,” *Educational Process: International Journal*, vol. 16, 2025, doi: 10.22521/edupij.2025.16.264.

[27] E. Yusliani, “Jurnal Penelitian dan Pengembangan Pendidikan Fisika,” *JPPPF*, vol. 8, no. 1, 2022, doi: 10.21009/1.

[28] S. Aryani, H. Novia, and A. Setiawan, “Development and Analysis Validation Instrument Critical Thinking Heat Concept (CTHT) Using Rash Model,” *Jurnal Penelitian Pendidikan IPA*, vol. 10, no. 7, pp. 3731–3738, Jul. 2024, doi: 10.29303/jppipa.v10i7.7727.

[29] F. D. P. Nesri and Y. D. Kristanto, “Pengembangan Modul Ajar Berbantuan Teknologi untuk Mengembangkan Kecakapan Abad 21 Siswa,” *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, vol. 9, no. 3, p. 480, Sep. 2020, doi: 10.24127/ajpm.v9i3.2925.

[30] Saheri, K. I. Supardi, and S. Haryani, “Pengembangan Instrumen Penilaian Keterampilan Berpikir Kritis Siswa SMA Melalui Model Pembelajaran Berbasis Masalah Materi Larutan Penyangga,” Aug. 2017.

[31] I. G. E. Saputra, I. N. Jampel, and I. G. L. A. Parwata, “Pengembangan Instrumen Penilaian Kemampuan Berpikir Kritis Peserta Didik Kelas VIII SMP pada Materi Getaran dan Gelombang,” *Jurnal Pendidikan dan Pembelajaran Sains Indonesia*, vol. 5, Oct. 2022.

[32] M. Ulfah and N. Imam Bonjol Padang, “instrumen pengujian produk pembelajaran (pengujian validitas, praktikalitas, efektivitas),” 2025.

[33] M. Safihin, H. S. Tiur Maria, and Hamdani, “pengembangan tes menggunakan model rasch materi gaya untuk SMA,” 2019.

[34] F. Nyimbili and L. Nyimbili, “Types of Purposive Sampling Techniques with Their Examples and Application in Qualitative Research Studies,” *British Journal of Multidisciplinary and Advanced Studies*, vol. 5, no. 1, pp. 90–99, Feb. 2024, doi: 10.37745/bjmas.2022.0419.

[35] R. H. Ennis, “Critical thinking: Its nature, measurement, and improvement.,” 1985.

[36] S. Yustina Purnama and S. Sri Wulandari, “Pengembangan Modul Administrasi Perkantoran pada Mata Pelajaran Administrasi Humas dan Keprotokolan Kelas XI APK di SMK Muhammadiyah 1 Lamongan,” 2021. [Online]. Available: https://journal.unesa.ac.id/index.php/jpap

[37] Y. B. Bhakti, R. Arthur, and Y. Supriyati, “The Content Validity: Multiple Choices Complex Physics

Instrument to Measure Critical Thinking Skills," *JIPF (Jurnal Ilmu Pendidikan Fisika)*, vol. 9, no. 1, p. 56, Feb. 2024, doi: 10.26737/jipf.v9i1.4765.

[38] R. P. Z, R. Sari, J. Jumadi, and A. Ariswan, "Pengembangan dan Validasi Instrumen Tes untuk Mengukur Keterampilan Menyelesaikan Masalah Peserta Didik SMA pada Pelajaran Fisika," *Jurnal Penelitian Pembelajaran Fisika*, vol. 11, no. 1, pp. 17–26, Mar. 2020, doi: 10.26877/jp2f.v11i1.3993.

[39] L. Missa, R. N. K. Mellu, and D. Langtang, "Pengembangan Alat Evaluasi Fisika Untuk Mengidentifikasi Miskonsepsi Siswa Kelas VIII SMP Negeri Oenino," *Jurnal Fisika Indonesia*, vol. 24, no. 3, p. 164, Dec. 2020, doi: 10.22146/jfi.v24i3.57639.

[40] N. Mappalesye, S. S. Sari, and K. Arafah, "pengembangan instrumen tes kemampuan berpikir kritis dalam pembelajaran fisika," *Jurnal Sains dan Pendidikan Fisika (JSPF) Jilid*, vol. 17, no. 1, pp. 69–82, Apr. 2021.

[41] M. subair, S. Salmah Sari, and Jasruddin, "pengembangan instrumen tes hasil belajar fisika kelas xi sma. Negeri 1 pangkajene development of instrument physics learning results in class xi science senior high school 1 in Pangkajene 1)," *Jurnal Sains dan Pendidikan Fisika*, vol. 15, no. 1, pp. 8–12, Apr. 2019, [Online]. Available: http://ojs.unm.ac.id

[42] D. Wahyuni, "Pengembangan Instrumen Tes Berpikir Kritis Mahasiswa Universitas Dharmawangsa," 2024.

[43] J. Jamaluddin, A. W. Jufri, M. Muhlis, and I. Bachtiar, "Pengembangan Instrumen Keterampilan Berpikir Kritis Pada Pembelajaran IPA di SMP," *Jurnal Pijar Mipa*, vol. 15, no. 1, pp. 13–19, Jan. 2020, doi: 10.29303/jpm.v15i1.1296.

[44] Y. H. Putri Hapsari and D. Gularso, "Pengembangan Kepraktisan Instrumen Penilaian Berbasis Keterampilan Berpikir Kritis Siswa Kelas IV SD Negeri Wirosaban," *Jurnal Cendekia : Jurnal Pendidikan Matematika*, vol. 8, no. 2, pp. 1351–1357, Jul. 2024, doi: 10.31004/cendekia.v8i2.3309.

[45] N. Siti Ratna Gumilang, Wahidin, and A. Tsurayya, "Pengembangan Instrumen Kemampuan Berpikir Kritis dan Kreatif Matematika Peserta Didik Kelas VII SMP Development of Mathematics' Critical and Creative Thinking Instruments for Grade VII on Secondary School," *Jurnal Pendidikan Matematika dan Sains*, vol. 9, no. 2, pp. 89–98, 2021, doi: 10.21831/jpms.v9i1.27349.

[46] Bahraeni Tajri Zahrah Via Faradisa, Septi Kurniasih, and Liska Berlian, "Pengembangan Instrumen Tes 4TMC CBT Pada Materi Sistem Pernapasan untuk Mengukur Berpikir Kritis Siswa SMP Kelas VIII," *JURNAL PENDIDIKAN MIPA*, vol. 14, no. 4, pp. 909–918, Dec. 2024, doi: 10.37630/jpm.v14i4.2010.

[47] N. N. Simamora, E. Istiyono, and I. Wilujeng, "Development of an Assessment Instrument to Measure High School Students' Collaboration Skills in Physics Learning," *JPPPF (Jurnal Penelitian dan Pengembangan Pendidikan Fisika)*, vol. 10, no. 1, Jun. 2024, doi: doi.org/10.21009/1.10104.

[48] K. Khotimah and Q. Fariyani, "Development of a Physics Assessment Instrument to Measure High School Students Multirepresentation Skills in Momentum and Impulse," *Impulse: Journal of Research and Innovation in Physics Education*, vol. 5, no. 1, pp. 49–61, Oct. 2025, doi: 10.14421/impulse.2025.51-04.

[49] P. W. Pradana, F. Febriani, M. Ibnusaputra, and J. Jumadi, "Development of Physics Test Instrument to Measure Verbal Representation of High School Student on Optical Instrument Topic," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 10, pp. 7963–7968, Oct. 2023, doi: 10.29303/jppipa.v9i10.3775.

[50] R. Sheptian, R. Arthur, and R. Wardhani, "psychometric analysis of mathematics readiness assessment tool for elementary school students: a rasch model approach," *Jurnal Inovasi Penelitian Pendidikan dan Pembelajaran*, vol. 4, no. 4, Nov. 2024, [Online]. Available: https://jurnalp4i.com/index.php/learning

[51] S. Nuryanti, M. Masykuri, and E. Susilowati, "Analisis Iteman dan model Rasch pada pengembangan instrumen kemampuan berpikir kritis peserta didik sekolah menengah kejuruan," *Jurnal Inovasi Pendidikan IPA*, vol. 4, no. 2, pp. 224–233, Oct. 2018, doi: 10.21831/jipi.v4i2.21442.

[52] S. D. S. Pradana, P. Parno, and S. K. Handayanto, "Pengembangan tes kemampuan berpikir kritis pada materi Optik Geometri untuk mahasiswa Fisika," *Jurnal Penelitian dan Evaluasi Pendidikan*, vol. 21, no. 1, pp. 51–64, Jun. 2017, doi: 10.21831/pep.v21i1.13139.

[53] R. H. Ennis, *A Logical Basis for Measuring Critical Thinking Skills.* 1985.

[54] E. Almanasreh, R. Moles, and T. F. Chen, "Evaluation of methods used for estimating content validity," *Research in Social and Administrative Pharmacy*, vol. 15, no. 2, pp. 214–221, Feb. 2019, doi: 10.1016/J.SAPHARM.2018.03.066.

[55] A. H. Aminudin, D. Rusdiana, A. Samsudin, L. Hasanah, and J. Maknun, "Measuring critical thinking skills of 11th grade students on temperature and heat," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1742-6596/1280/5/052062.

[56] E. Avinç and F. Doğan, "Digital literacy scale: Validity and reliability study with the rasch model," *Educ. Inf. Technol. (Dordr).*, vol. 29, no. 17, pp. 22895–22941, Dec. 2024, doi: 10.1007/s10639-024-12662-7.