EDUTECH : Jurnal Inovasi Pendidikan Berbantuan Teknologi
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2797-0140 | p-ISSN : 2797-0590
Online Journal System : https://jurnalp4i.com/index.php/edutech

STEM-BASED DIFFERENTIATION ASSESSMENT TOOLS FOR CRITICAL THINGKING SKILLS IN TRANSVERSAL LEARNING

Melly Abdellina Siboro[1], Junistel Rajagukguk[2]
Medan State University[12]
e-mail: juniastel@unimed.ac.id

Diterima: *21/07/2026*; Direvisi: *03/08/2026*; Diterbitkan: *07/08/2026*

ABSTRAK

Berpikir kritis merupakan salah satu kompetensi penting abad ke-21 yang harus dimiliki peserta didik. Namun, pembelajaran fisika di sekolah masih didominasi oleh soal-soal rutin yang hanya mengukur kemampuan mengingat dan pemahaman konsep sehingga potensi peserta didik dalam berpikir tingkat tinggi belum terukur secara optimal. Penelitian dan pengembangan (*Research and Development*/R&D) ini bertujuan mengembangkan instrumen asesmen berdiferensiasi berbasis STEM (*Science, Technology, Engineering, and Mathematics*) yang valid, reliabel, praktis, dan efektif untuk mengukur kemampuan berpikir kritis peserta didik SMA pada materi gelombang transversal. Penelitian menggunakan model pengembangan 4D (*Define, Design, Develop,* dan *Disseminate*). Subjek penelitian terdiri atas tiga validator ahli, seorang guru fisika, dan 20 peserta didik kelas XI MIPA di SMAS YAPIM TARUNA STABAT. Data dikumpulkan melalui lembar validasi ahli, tes kemampuan berpikir kritis, serta angket respons guru dan peserta didik. Analisis data meliputi uji validitas dan reliabilitas, analisis tingkat kesukaran dan daya pembeda butir soal, uji Wilcoxon, serta analisis N-Gain. Hasil penelitian menunjukkan bahwa instrumen memperoleh tingkat validitas ahli sebesar 88% (sangat valid) dengan koefisien Cohen's Kappa sebesar 0,590 (kategori baik). Empat dari enam butir soal dinyatakan valid secara empiris, sedangkan reliabilitas instrumen meningkat dari 0,210 menjadi 0,793 setelah dilakukan penyempurnaan butir soal. Rata-rata skor kemampuan berpikir kritis peserta didik meningkat secara signifikan dari 34,95 pada pretest menjadi 93,52 pada posttest, dengan nilai signifikansi uji Wilcoxon sebesar p = 0,000 (< 0,05) dan nilai N-Gain sebesar 0,8977 (kategori tinggi). Hasil uji kepraktisan menunjukkan persentase sebesar 94,32% (sangat praktis). Temuan ini menunjukkan bahwa instrumen yang dikembangkan telah memenuhi kriteria valid, reliabel, praktis, dan efektif sehingga layak digunakan sebagai alternatif instrumen asesmen untuk mengukur kemampuan berpikir kritis dalam pembelajaran fisika abad ke-21 yang kontekstual, berkeadilan, dan bermakna bagi peserta didik dengan karakteristik yang beragam.

Kata kunci: *instrumen asesmen, asesmen berdiferensiasi, STEM, berpikir kritis, gelombang transversal*

ABSTRACT

Critical thinking is an essential 21st-century competency that students are expected to develop. However, physics instruction in schools remains largely focused on routine problems that primarily assess memory and conceptual understanding, limiting opportunities to evaluate higher-order thinking skills. This research and development (R&D) study aimed to develop a STEM (Science, Technology, Engineering, and Mathematics)-based differentiated assessment instrument that is valid, reliable, practical, and effective for measuring senior high school students' critical thinking skills on the topic of transverse waves. The study employed the 4D development model (Define, Design, Develop, and Disseminate). The participants consisted of

https://doi.org/10.51878/edutech.v6i3.14299

three expert validators, one physics teacher, and 20 Grade 11 science students at SMAS YAPIM TARUNA STABAT. Data were collected using expert validation sheets, critical thinking tests, and teacher and student response questionnaires. Data analysis included validity and reliability testing, item difficulty and discrimination analyses, the Wilcoxon signed-rank test, and N-Gain analysis. The results indicated that the instrument achieved an expert validity score of 88% (highly valid) with a Cohen's Kappa coefficient of 0.590 (good agreement). Four of the six test items were empirically valid, and the instrument's reliability improved from 0.210 to 0.793 after item refinement. Students' mean critical thinking scores increased substantially from 34.95 on the pretest to 93.52 on the posttest, with a Wilcoxon significance value of $p = 0.000$ (< 0.05) and an N-Gain score of 0.8977 (high category). The practicality evaluation yielded a score of 94.32% (highly practical). These findings demonstrate that the developed instrument is valid, reliable, practical, and effective for assessing critical thinking skills and can serve as an alternative assessment tool to support contextual, equitable, and meaningful physics learning in the 21st century.
Keywords: *assessment instrument, differentiation, STEM, critical thinking, transverse waves*

INTRODUCTION

Critical thinking is one of the essential competencies required in twenty-first-century education because it enables learners to analyze information, evaluate evidence, make reasoned decisions, and solve complex real-world problems (Saputra et al., 2022; Permata et al., 2025). In physics education, critical thinking plays a particularly important role because many physical phenomena require students to connect conceptual understanding, mathematical reasoning, and scientific evidence rather than merely recalling formulas. Consequently, physics learning should be accompanied by assessment instruments capable of measuring students' higher-order thinking skills in authentic contexts.

Despite its importance, physics assessment practices in many Indonesian schools remain dominated by routine exercises that emphasize factual recall and algorithmic problem solving. This condition was also identified at SMAS YAPIM TARUNA STABAT, where teachers reported difficulties in developing assessment instruments that simultaneously measure critical thinking and integrate authentic STEM (Science, Technology, Engineering, and Mathematics) contexts. As a result, students have limited opportunities to demonstrate analytical reasoning, evaluate alternative solutions, and apply scientific concepts to real-world situations. This challenge is reflected in Indonesia's performance in the Programme for International Student Assessment (PISA). In 2018, Indonesian students achieved a science literacy score of only 396, declining from 403 in 2015 and remaining substantially below the OECD average of 487 (Hapsari & Gularso, 2024). Low science literacy indicates students' limited ability to interpret scientific evidence, construct logical arguments, and solve complex problems, all of which represent core dimensions of critical thinking.

Another challenge concerns the use of standardized assessment instruments that apply identical questions to students with diverse academic readiness levels. Such assessments frequently generate learning anxiety among lower-achieving students while providing insufficient cognitive challenge for higher-achieving students (Idamayanti & Sakti, 2025; Supriana et al., 2024). Differentiated assessment has therefore been recommended as an alternative approach because it adapts assessment tasks to students' readiness, interests, and learning profiles while maintaining equivalent learning objectives (Tomlinson, 2013; Maharani et al., 2024; Marlina et al., 2023). Through differentiated assessment, every student is provided with equitable opportunities to demonstrate higher-order thinking according to their learning

https://doi.org/10.51878/edutech.v6i3.14299

characteristics. Nevertheless, practical implementation of differentiated assessment in physics, particularly within critical thinking assessment instruments, remains limited.

The STEM approach provides a promising framework for developing critical thinking because each discipline contributes complementary reasoning processes. Science encourages evidence-based inquiry; Technology emphasizes the effective utilization of digital resources; Engineering develops systematic problem solving through design and optimization; and Mathematics strengthens logical reasoning and quantitative analysis (Kusyanto et al., 2022; Amelia Purnomo & Puspitawati, 2025; Arjaya et al., 2025). Previous studies have demonstrated that STEM-based learning materials, worksheets, and augmented reality media can improve students' critical thinking skills (Hadiyanti et al., 2021; Oktaviyanti et al., 2023; Arisha & Surya, 2024). However, improvements in learning cannot be accurately demonstrated without assessment instruments specifically designed to measure critical thinking within integrated STEM contexts.

Among physics topics taught in senior high school, transverse waves require students to connect abstract mathematical representations with observable physical phenomena, such as vibrations on strings, wave propagation, wavelength, frequency, and amplitude. These concepts demand analytical reasoning, interpretation of multiple representations, and application of scientific principles rather than procedural calculations alone. Consequently, transverse waves provide an appropriate context for assessing higher-order thinking skills through authentic STEM-based problems.

Several studies have developed higher-order thinking or critical thinking assessment instruments in various disciplines and physics topics. Existing instruments have been developed for biology (Abeto, 2024; Hasibuan et al., 2024), vibrations and waves (Saputra et al., 2022), sound waves (Khoiriyah et al., 2018), light waves (Kumalasari & Putra, 2021), running waves (Priatna et al., 2024), and standing and running waves (Setiani et al., 2023). Other studies have focused on differentiated assessment (Maharani et al., 2024; Marlina et al., 2023) or STEM-based learning interventions (Amelia Purnomo & Puspitawati, 2025; Arisha & Surya, 2024). Nevertheless, no previous study has developed a critical thinking assessment instrument that simultaneously integrates STEM principles and differentiated assessment specifically for the topic of transverse waves in senior high school physics.

Accordingly, this study contributes to the literature in three ways. First, it develops a differentiated assessment instrument that integrates all STEM components into authentic critical thinking tasks. Second, the instrument is specifically designed for the topic of transverse waves, an area that has received limited attention in previous instrument development studies. Third, the study comprehensively evaluates the instrument through expert validation, empirical validity, reliability, discrimination index, difficulty level, user responses, and effectiveness using the Four-D (4-D) research and development model proposed by Thiagarajan (1974). Therefore, this study aims to develop and evaluate a STEM-based differentiated assessment instrument for measuring the critical thinking skills of twelfth-grade senior high school students on the topic of transverse waves. The findings are expected to provide a valid, reliable, and practical assessment instrument that supports differentiated physics learning while contributing to the advancement of STEM-oriented assessment practices.

RESEARCH METHODOLOGY

This study employed a Research and Development (R&D) approach using the Four-D (4-D) development model proposed by Thiagarajan (1974), which consists of four stages: Define, Design, Develop, and Disseminate. The model was selected because it provides a

https://doi.org/10.51878/edutech.v6i3.14299

systematic framework for developing educational products and evaluating their quality before implementation. The research was conducted at SMAS YAPIM TARUNA STABAT, Langkat Regency, North Sumatra, Indonesia, from March to April 2026.

The research participants consisted of **20 eleventh-grade science students** who participated in the limited field trial of the developed instrument. Prior to the field trial, the assessment instrument was evaluated by three expert validators comprising two university lecturers in physics education and one experienced senior high school physics teacher. The validators assessed the instrument using a binary-scale validation sheet (appropriate/inappropriate) covering six aspects: (1) content validity, (2) construct validity, (3) language clarity, (4) STEM integration, (5) differentiated assessment characteristics, and (6) critical thinking components.

The product developed in this study was a STEM-based differentiated assessment instrument consisting of six open-ended essay questions. The items were constructed based on Ennis's critical thinking framework, including elementary clarification, basic support, inference, advanced clarification, and strategy and tactics, and were designed to measure higher-order cognitive processes at Bloom's revised taxonomy levels C4 (analyzing), C5 (evaluating), and C6 (creating). Authentic STEM contexts were incorporated into the assessment through real-life situations involving vibrating guitar strings, suspension bridge cables, polarized light, and the PhET *Wave on a String* simulation.

Differentiated assessment principles were embedded into the instrument through three dimensions. Content differentiation was implemented by providing questions with varying levels of complexity according to students' readiness. Process differentiation allowed students to complete tasks either through independent analytical work or by utilizing interactive PhET simulations. Product differentiation enabled students to demonstrate their understanding through written explanations, diagrams, mathematical analyses, or simulation reports. Despite these variations, all student responses were evaluated using a single analytical scoring rubric to ensure consistency and fairness.

The data were analyzed using IBM SPSS Statistics. Content validity was determined through the percentage of validator agreement, while inter-rater reliability was examined using Cohen's Kappa coefficient. Empirical validity of each item was analyzed using Pearson's product-moment correlation, whereas internal consistency reliability was measured using Cronbach's Alpha coefficient. Item quality was further evaluated through difficulty index and discrimination index analyses by comparing the upper and lower groups, each representing 27% of the participants. Before evaluating the effectiveness of the instrument, pretest and posttest data were tested for normality using the Shapiro–Wilk test. Because the posttest data were not normally distributed, differences between pretest and posttest scores were analyzed using the Wilcoxon Signed-Rank Test at a significance level of 0.05. Furthermore, improvements in students' critical thinking skills were quantified using the normalized gain (N-Gain) index. The practicality of the developed instrument was evaluated through teacher and student response questionnaires, and the results were analyzed descriptively using percentage scores.

RESULTS AND DISCUSSION

The development results show that the instrument successfully passed through all stages of the 4D model, from needs analysis to limited field testing. This section presents the main findings in sequence: expert validation results, empirical test results (validity, reliability, difficulty level, discriminant power), effectiveness test results, and practicality test results, followed by a discussion interpreting these findings.

https://doi.org/10.51878/edutech.v6i3.14299

EDUTECH : Jurnal Inovasi Pendidikan Berbantuan Teknologi
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2797-0140 | p-ISSN : 2797-0590
Online Journal System : https://jurnalp4i.com/index.php/edutech

Results

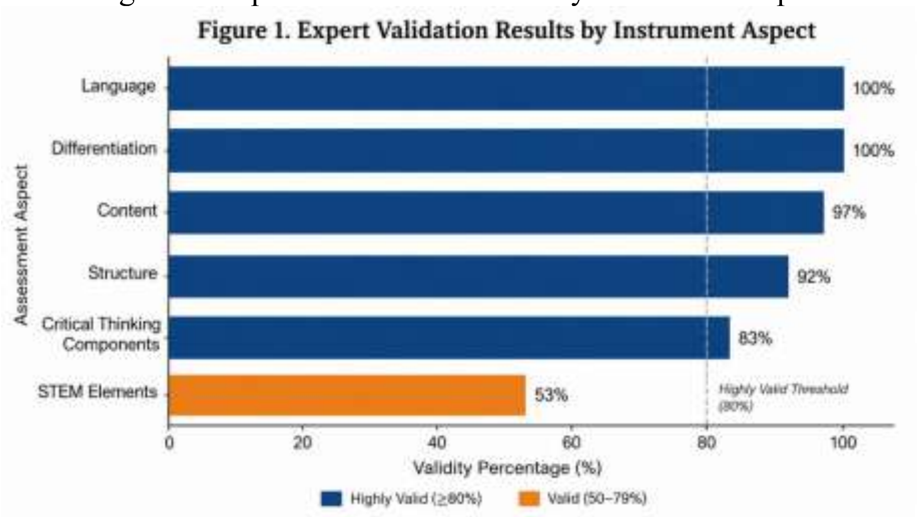
Expert Validation

The instrument was evaluated by three expert validators across six aspects using a binary scale (appropriate = 1, inappropriate = 0). Table 1 and Figure 1 present the complete results.

Table 1. Expert Validation Results by Instrument Aspect

No.	Assessment Aspect	Maximum Score	Validator 1 & 2	Total Score	Percentage	Category
1	Content Validity	18	17 & 18	35	97%	Highly Valid
2	Structure	18	17 & 16	33	92%	Highly Valid
3	Language	18	18 & 18	36	100%	Highly Valid
4	STEM Elements	18	9 & 10	19	53%	Valid
5	Differentiation	18	18 & 18	36	100%	Highly Valid
6	Critical Thinking Components	12	10 & 10	20	83%	Highly Valid
Total	**Overall**	**102**	**89 & 90**	**179**	**88%**	**Highly Valid**

Figure 1. Expert Validation Results by Instrument Aspect



Based on Table 1 and Figure 1, the language and differentiation aspects achieved the highest scores (100%, highly valid), followed by content (97%), structure (92%), and critical thinking (83%). Only the STEM elements fell into the "valid" category (53%), not because the test items were poorly written, but because the Technology and Engineering components were present only as a scenario framework, rather than as actual design or testing activities performed by students, due to limited access to laboratories and a lack of basic engineering experience in schools (Hadiyanti et al., 2021; Srirahayu & Arty, 2018). Overall, the instrument's validity reached 88%, categorized as highly valid. Inter-rater agreement was tested using Cohen's

https://doi.org/10.51878/edutech.v6i3.14299

Kappa, yielding a value of 0.590 (good category, significance 0.000), indicating that the two raters' assessments converged on most criteria, though they were not entirely identical.

Item-by-item analysis revealed some interesting variations: Item 1 obtained a Kappa of 0.821 (very good) and Item 2 achieved perfect agreement among validators, while Item 3 obtained a Kappa of −0.097 (low) because one of the validators assigned scores that showed little variation, a finding that is consistent with the results of the empirical validity test for Item 3, which was also deemed invalid.

Empirical Validity, Reliability, Difficulty Level, and Discriminant Power

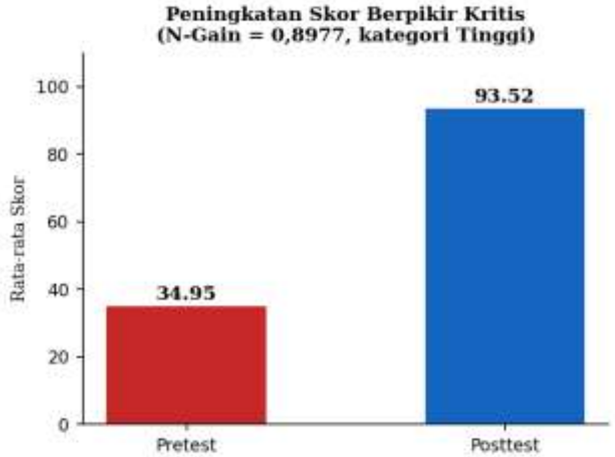
The field test involved 20 students. The results of the empirical validity test (Pearson correlation) showed that four of the six items, namely, Item 2 ($r = 0.518$), Item 4 ($r = 0.736$), Item 5 ($r = 0.790$), and Item 6 ($r = 0.650$), were deemed valid ($p < 0.05$), while Item 1 ($r = 0.114$) and Item 3 ($r = -0.176$) were deemed invalid. Before these two items were removed, the instrument's reliability (Cronbach's Alpha) was only 0.210 (low); after item selection, reliability jumped to 0.793 (good). This sharp increase is consistent with the finding that Items 1 and 3 exhibited a negative discriminant power pattern, that is, some low-ability students answered correctly more often than high-ability students, thereby contributing to random variance that reduced the consistency of the total score.

An analysis of difficulty levels showed that Items 1, 2, and 4 fell into the moderate category, while Items 3, 5, and 6 were classified as difficult, in line with the instrument's Higher-Order Thinking Skills (HOTS) orientation. However, high difficulty is not the sole cause of invalidity, as Questions 5 and 6 are both difficult yet remain valid; the issues with Questions 1 and 3 are more related to ambiguous stimuli and unclear question wording. The discriminant validity of the four valid questions ranges from adequate to good, indicating that these questions are able to consistently distinguish between high- and low-ability students.

Effectiveness of the Instrument

After being revised based on the validators' feedback, the instrument was pilot-tested to measure its effectiveness in improving students' critical thinking skills using a one-group pretest-posttest design. The average student score increased sharply from 34.95 (pretest) to 93.52 (posttest), with the score distribution also narrowing (the standard deviation decreased from 10.346 to 5.183), indicating a consistent improvement across all students, not just among a select few (Figure 2).

Figure 2. Comparison of Average Pretest and Posttest Scores



https://doi.org/10.51878/edutech.v6i3.14299

The Shapiro–Wilk normality test showed that the pretest scores were normally distributed (Sig. = 0.113 > 0.05) but the posttest scores were not normally distributed (Sig. = 0.042 < 0.05); therefore, the hypothesis testing proceeded with the Wilcoxon Signed-Rank Test (nonparametric). The results showed a p-value of 0.000 (< 0.05), indicating a statistically significant difference between the scores before and after the implementation of the instrument. The normalized N-Gain test reinforced these findings with a mean of 0.8977, which falls into the high category (range 0.72–1.00).

Practicality of the Instrument

A practicality test conducted via a response questionnaire showed that students rated the material at 88.64% (very practical), primarily due to the clear language and instructions as well as the relevant real-life context, while teachers rated it at 100% (very practical) due to the systematic question structure and objective assessment rubric. The combined average practicality score was 94.32%, categorized as very practical (Table 2).

Table 2. Summary of the Instrument's Practicality Test Results

Respondent	Percentage	Category
Students	88.64%	Highly Practical
Teacher	100%	Highly Practical
Average	**94.32%**	**Highly Practical**

Discussion

The four findings above, expert validity, empirical validity and reliability, effectiveness, and practicality, essentially point to a common thread: the strongest aspect of the instrument is its flexible differentiation design and the high-level critical thinking it embodies, while its STEM comprehensiveness, particularly the Technology and Engineering components, is an aspect that has only been partially fulfilled.

The development procedure, which progresses from needs analysis, item development based on Ennis's indicators and the STEM context, expert validation and revision, to empirical testing and effectiveness assessment, aligns with differentiation assessment theory, which emphasizes that instruments must be designed flexibly from the outset, rather than adapted later (Tomlinson, 2013; Chapman & King, 2012). The fact that the expert validation stage yielded a high score (88%) but the empirical stage still left two weak items is not a contradiction; rather, it is evidence that the two-stage validation design worked as intended, namely, to filter out weaknesses not detected theoretically. A similar pattern was found in the study by Setiani et al. (2023), whose instrument also lost a number of items during the empirical stage despite its strong theoretical design, indicating that the elimination of some items between the expert and empirical stages is a common pattern in wave-themed critical thinking instruments, rather than a specific weakness of this research procedure.

The low score for the STEM elements (53%) should be interpreted in context, rather than generalized as a weakness of the instrument as a whole. Science and Mathematics were activities that students actually engaged in, such as calculating and analyzing simulation data, whereas Technology and Engineering served only as the framework for the story scenarios, meaning students did not actually design or test anything. Conversely, the differentiation element was actually the strongest aspect, validated alongside language (both at 100%), indicating that the designed content, processes, and products were deemed fully appropriate for

https://doi.org/10.51878/edutech.v6i3.14299

EDUTECH : Jurnal Inovasi Pendidikan Berbantuan Teknologi
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2797-0140 | p-ISSN : 2797-0590
Online Journal System : https://jurnalp4i.com/index.php/edutech

accommodating the diversity of students' readiness, interests, and learning profiles. Shortcomings in the STEM aspect should by no means be interpreted as weaknesses in the instrument's differentiation or critical thinking aspects, which are in fact its main strengths.

The jump in reliability from 0.210 to 0.793 after removing the two weak items has a specific explanation: both items exhibited negative discriminant validity, thereby contributing to random variance that is actually treated as "noise" by the Alpha formula (Dessiane & Kristin, 2021; Utami & Wardani, 2020). It is important to note that Items 1 and 3 represent two of Ennis's more basic indicators (Elementary Clarification and Basic Support), so all reliability and validity figures reported thereafter actually describe a four-item instrument weighted toward the higher-level indicators (Inference, Advanced Clarification, and Strategy and Tactics), rather than a measurement that is evenly distributed across all five Ennis indicators. Compared to the ongoing-wave HOTS instrument by Priatna et al. (2024), which achieved a reliability of 0.90 with items within a narrower difficulty range, the reliability obtained here is most likely due to the inclusion of two difficult-ambiguous items in the empirical stage, a concrete point to address in future revisions.

The high effectiveness of the instrument (N-Gain 0.8977) aligns with the argument that integrated STEM learning naturally guides students toward the levels of analyzing, evaluating, and creating, as each component of STEM requires a different type of reasoning (Amelia Purnomo & Puspitawati, 2025; Kusyanto et al., 2022; Arjaya et al., 2025). This N-Gain value is even higher than that of the STEM-based worksheets by Arisha and Surya (2024), which achieved an N-Gain of 0.71, and far exceeds that of the STEM-AR instructional materials by Oktaviyanti et al. (2023), which only reached the moderate category, likely because this instrument was explicitly developed based on Ennis's indicators with differentiated response options, so the intervention focused directly on the constructs being measured, rather than merely general exposure to STEM.

Meanwhile, the gap between students' practicality scores (88.64%) and teachers' scores (100%) does not mean that teachers rated the instrument better than students; rather, it is because the two groups assessed different aspects: students evaluated ease of use and understanding of wave content, while teachers evaluated ease of administration and the objectivity of the rubric. These findings are consistent with those of Maharani et al. (2024) and Marlina et al. (2023), who reported that differentiated assessment instruments tend to be well-received by users even though the differentiation construct itself is still being empirically refined.

As with most development research, this study has limitations: the pilot test was conducted at only one school with 20 students, so the generalizability of the results remains limited; two test items still require further revision; the integration of technology and engineering has not been fully realized due to limited laboratory facilities; the duration of the pilot study was relatively short; and the scope of the research is still limited to the topic of transverse waves and critical thinking skills alone, not yet addressing other 21st-century skills such as creativity, communication, and collaboration.

CONCLUSION

A STEM-based differentiated assessment instrument for the topic of transverse waves was successfully developed using the Four-D (4-D) development model and demonstrated satisfactory validity, reliability, effectiveness, and practicality. Expert validation yielded an overall validity score of 88% (highly valid) with good inter-rater agreement (Cohen's Kappa = 0.590). Four of the six essay items were empirically valid, and the instrument's reliability

https://doi.org/10.51878/edutech.v6i3.14299

EDUTECH : Jurnal Inovasi Pendidikan Berbantuan Teknologi
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2797-0140 | p-ISSN : 2797-0590
Online Journal System : https://jurnalp4i.com/index.php/edutech

improved substantially from 0.210 to 0.793 following item refinement. The instrument was effective in enhancing students' critical thinking skills, as indicated by an increase in the mean score from 34.95 on the pretest to 93.52 on the posttest, a statistically significant Wilcoxon Signed-Rank Test result ($p < 0.001$), and an N-Gain score of 0.8977 (high category). The practicality evaluation also demonstrated highly positive results, with an overall practicality score of 94.32% based on teacher and student responses. These findings indicate that the developed instrument is suitable for assessing critical thinking skills in contextual, equitable, and meaningful 21st-century physics learning. Nevertheless, the STEM Technology and Engineering components require further refinement to strengthen the overall quality of the instrument. Future studies are recommended to involve larger and more diverse samples across multiple schools, revise the two empirically invalid items, enhance the Technology and Engineering dimensions, and evaluate the applicability of the instrument to other physics topics.

REFERENCES

Abeto, A. (2024). Pengembangan instrumen penilaian keterampilan berpikir tingkat tinggi untuk kelas XI. *Jurnal Pendidikan Tambusai*, 8(1), 6785–6799. https://doi.org/10.31004/jptam.v8i1.13437

Amelia Purnomo, P., & Puspitawati, R. P. (2025). Pengembangan e-modul berbasis STEM pada materi pencemaran udara untuk melatihkan keterampilan berpikir kritis siswa. *BioEdu: Berkala Ilmiah Pendidikan Biologi*, 14(3), 630–638. https://doi.org/10.26740/bioedu.v14n3.p630-638

Arisha, D., & Surya, E. (2024). Pengembangan LKPD berbasis pendekatan STEM (Science, Technology, Engineering, and Mathematics) untuk meningkatkan kemampuan berpikir kritis. *Jurnal Review Pendidikan dan Pengajaran*, 7(1), 345–352. https://doi.org/10.31004/jrpp.v7i1.24172

Arjaya, B. A. I., Puspaningsih, R. A., Ekayanti, N. W., Budiningsih, N. D., & Suyati. (2025). Optimalisasi kemampuan berpikir kritis siswa melalui implementasi pendekatan STEM di SMAN Bali Mandara. *Jurnal Santiaji Pendidikan*, 15(1), 112–121. https://doi.org/10.36733/jsp.v15i1.11127

Dessiane, S. T., & Kristin, F. (2021). Pengembangan instrumen penilaian sikap sosial pembelajaran tematik kelas IV SD. *Jurnal Pendidikan Ilmu Pengetahuan Sosial Indonesia*, 6(1), 21–26. https://doi.org/10.26737/jpipsi.v6i1.2310

Hadiyanti, N. F. D., Hobri, Prihandoko, A. C., Susanto, Murtikusuma, R. P., Khasanah, N., & Maharani, P. (2021). Development of mathematics e-module with STEM-collaborative project-based learning to improve mathematical literacy ability of vocational high school students. *Journal of Physics: Conference Series*, 1839(1). https://doi.org/10.1088/1742-6596/1839/1/012031

Hapsari, Y. H. P., & Gularso, D. (2024). Pengembangan kepraktisan instrumen penilaian berbasis keterampilan berpikir kritis siswa kelas IV SD Negeri Wirosaban. *Jurnal Cendekia: Jurnal Pendidikan Matematika*, 8(2), 1351–1357. https://doi.org/10.31004/cendekia.v8i2.3309

Hasibuan, Y. V., Hasairin, A., & Restuati, M. (2024). Pengembangan instrumen penilaian hasil belajar bioteknologi kelas XII SMA berbasis HOTS dan kemampuan berpikir kritis. *Didaktika: Jurnal Kependidikan*, 13(4), 4317–4326. https://doi.org/10.58230/27454312.1092

https://doi.org/10.51878/edutech.v6i3.14299

EDUTECH : Jurnal Inovasi Pendidikan Berbantuan Teknologi
Vol. 6, No. 3, Juni-Agustus 2026
e-ISSN : 2797-0140 | p-ISSN : 2797-0590
Online Journal System : https://jurnalp4i.com/index.php/edutech

Idamayanti, R., & Sakti, I. (2025). Pengembangan modul ajar fisika berdiferensiasi di SMA. *Charm Sains: Jurnal Pendidikan Fisika*, 6(1), 17–24. https://doi.org/10.53682/charmsains.v6i1.416

Kumalasari, S. N., & Putra, N. M. D. (2021). Pengembangan instrumen penilaian keterampilan berpikir kritis pada materi gelombang cahaya untuk siswa SMA. *UPEJ Unnes Physics Education Journal*, 10(3), 256–263. https://doi.org/10.15294/upej.v10i3.55696

Kusyanto, K., Irwan, E., & Yazid, I. (2022). Implementasi pendekatan STEM untuk meningkatkan kemampuan berpikir kritis, berpikir kreatif, dan self-efficacy. *Pasundan Journal of Mathematics Education*, 12(2), 1–16. https://doi.org/10.23969/pjme.v12i2.5438

Maharani, P. R., Nitiasih, P. K., & Mahayanti, N. W. S. (2024). Differentiated assessment instrument: A development study (a case study of grade XI students in the first semester of senior high school). *Jurnal Review Pendidikan dan Pengajaran*, 7(4), 13928–13941. https://doi.org/10.31004/jrpp.v7i4.34784

Marlina, M., Kusumastuti, G., & Ediyanto, E. (2023). Differentiated learning assessment model to improve involvement of special needs students in inclusive schools. *International Journal of Instruction*, 16(4), 423–440. https://doi.org/10.29333/IJI.2023.16425A

Oktaviyanti, R., Fatmahanik, U., & Fadly, W. (2023). Pengembangan bahan ajar berbasis STEM dengan memanfaatkan augmented reality dalam meningkatkan kemampuan berpikir kritis. *Jurnal Tadris IPA Indonesia*, 3(3), 303–314. https://doi.org/10.21154/jtii.v3i3.1592

Permata, I., Nugroho, A. S., & Hayat, M. S. (2025). Pengembangan instrumen penilaian keterampilan berpikir kritis pada materi ekosistem untuk siswa kelas X: Validitas dan reliabilitas. *Eduproxima (Jurnal Ilmiah Pendidikan IPA)*, 7(4), 2053–2063. https://jurnal.stkippgritulungagung.ac.id/index.php/eduproxima/article/view/8107

Priatna, R., Agustina, T. W., & Malik, A. (2024). Pengembangan instrumen tes kemampuan berpikir tingkat tinggi pada materi gelombang berjalan. *Science and Physics Education Journal*, 8(1), 22–33. https://doi.org/10.31539/spej.v8i1.12628

Saputra, I. G. E., Jampel, I. N., & Parwata, I. G. L. A. (2022). Pengembangan instrumen penilaian kemampuan berpikir kritis peserta didik kelas VIII SMP pada materi getaran dan gelombang. *Jurnal Pendidikan dan Pembelajaran Sains Indonesia*, 5(2), 154–164. https://doi.org/10.23887/jppsi.v5i2.51772

Setiani, R., Kafii, M. S., & Dwikoranto. (2023). Validity development of critical thinking skill test instruments on running and stationary wave materials. *Inovasi Pendidikan Fisika*, 12(3), 111–118. https://doi.org/10.26740/ipf.v12n3.p111-118

Supriana, E., Liliani, N. T., & Luthfia, R. Z. (2024). Tantangan implementasi pembelajaran berdiferensiasi: Sebuah studi literatur. *Jurnal Pembelajaran, Bimbingan, dan Pengelolaan Pendidikan*, 4(5). https://doi.org/10.17977/um065.v4.i5.2024.9

Thiagarajan, S. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Indiana University. *(Buku, tidak memiliki DOI)*

Tomlinson, C. A. (2013). *Assessment and student success in a differentiated classroom*. ASCD.

Utami, D. A. P., & Wardani, N. S. (2020). Pengembangan instrumen penilaian kognitif dalam pembelajaran tematik kelas V SD. *Jurnal Ilmiah Kependidikan*, 20(2), 1–18. https://jurnal.stkippgribl.ac.id/index.php/lentera/article/view/483

https://doi.org/10.51878/edutech.v6i3.14299