

Development of Authentic Assessment Instrument to Measure Students Critical Thinking Skills on Temperature and Heat Subject Matter at SMAN 2 Medan

Lioni Tampubolon*, Yeni Megalina

Universitas Negeri Medan, Indonesia

ARTICLE INFO

Article history:

Submitted : July 10th, 2025

Revised : September 17th, 2025

Accepted : December 11th, 2025

Keywords:

Authentic assessment, Instrument assessment, Critical thinking, Temperature and heat



ABSTRACT

This research aims to develop an authentic assessment instrument in the form of cause-and-effect multiple-choice questions to measure students' critical thinking skills on the topic of temperature and heat at SMAN 2 Medan. The study is motivated by the limited implementation of authentic assessments in physics learning and the lack of students' engagement in context-based evaluations that foster critical thinking. The research employed a Research and Development (R&D) method using the ADDIE development model, which includes Analysis, Design, Development, Implementation, and Evaluation stages. Validation was conducted by two university lecturers and one physics teacher, followed by small and large group trials, and item analysis was performed to evaluate validity, reliability, difficulty level, discrimination index, and distractor effectiveness. Teacher responses were also gathered to assess the feasibility and practicality of the instrument. The results indicate that the developed instrument met validity and reliability criteria, with most items demonstrating appropriate levels of difficulty and discrimination. Teachers responded positively, stating that the instrument was clear, practical, and aligned with the needs of physics learning under the Merdeka Curriculum. Based on the findings of the study, it can be concluded that the authentic assessment instrument in the form of cause-and-effect items is feasible and effective in measuring students' critical thinking skills on the topic of temperature and heat. The implication is that this instrument can serve as an alternative evaluation tool for teachers to assess students' understanding in a deeper and more contextual manner while promoting active engagement in critical thinking.

© COPYRIGHT (C) 2025 PHYSICS EDUCATION RESEARCH JOURNAL

Introduction

Developing students' critical thinking skills has become one of the primary goals of science education in the twenty-first century. Rapid technological advancement, increasingly complex scientific problems, and the growing demand for evidence-based decision-making require students not only to master scientific concepts but also to analyze information critically, evaluate evidence, construct logical arguments, and solve authentic problems. Consequently, critical thinking has been recognized as one of the essential competencies for preparing students to face future educational, professional, and

societal challenges (Barevičiūtė et al., 2023; Daengneam et al., 2025; Shaber et al., 2025).

Physics education provides an ideal context for fostering critical thinking because students are continuously required to interpret phenomena, formulate hypotheses, analyze experimental evidence, and apply scientific reasoning to solve complex problems. However, developing these competencies cannot rely solely on instructional innovation; it also requires assessment systems capable of accurately measuring higher-order thinking processes. Recent studies emphasize that assessment should function not merely as a tool for measuring learning outcomes

but also as an integral component that supports students' cognitive development and guides instructional improvement (Arce-Trigatti et al., 2022; Jiang et al., 2023; Kumar et al., 2023).

Traditional paper-and-pencil assessments remain widely used in physics classrooms, yet they frequently emphasize factual recall and algorithmic problem solving rather than students' reasoning processes. Such assessments provide limited evidence regarding how students analyze scientific information, evaluate alternative explanations, or apply concepts to authentic situations. In contrast, authentic assessment engages students in meaningful tasks that resemble real-world contexts, thereby allowing teachers to evaluate both learning products and cognitive processes simultaneously. Previous studies have shown that authentic and performance-based assessments produce more comprehensive evidence of students' higher-order thinking, scientific reasoning, and problem-solving abilities than conventional testing approaches (Adri & Abdullah, 2022; Zdravkovska-Adamova et al., 2024).

The increasing emphasis on authentic assessment has stimulated extensive research on developing assessment instruments across educational disciplines. Researchers have designed instruments to assess critical thinking through performance assessment (Adri & Abdullah, 2022), formative assessment in physics (Pals et al., 2023), chemistry learning integrated with science process skills (Kartowagiran & Rohaeti, 2021), collaboration and responsibility skills in physics learning (Rosidin et al., 2022), graph representation ability in sound waves (Pradana, 2025), and measurement uncertainty in physics laboratories (Geschwind et al., 2024a, 2024b). Numerous studies have also highlighted the importance of establishing strong psychometric properties, including validity and reliability, during instrument development to ensure that assessment results accurately represent students' (Asriadi et al., 2026; Geschwind et al., 2024a).

Although substantial progress has been made in developing assessment instruments, several research gaps remain. First, many existing critical thinking assessments employ general problem-solving tasks without sufficiently embedding authentic contexts encountered by students during physics learning. Second, many instruments primarily measure cognitive outcomes using written tests and provide limited opportunities to evaluate students' reasoning processes during authentic scientific activities. Third, relatively few studies have developed authentic assessment instruments specifically designed for secondary-school physics topics that are conceptually challenging, such as temperature and heat.

Temperature and heat constitute one of the most fundamental yet conceptually difficult topics in secondary physics because students frequently experience misconceptions related to thermal equilibrium, heat transfer, specific heat capacity, and the distinction between heat and temperature. These concepts require students to interpret experimental evidence, connect multiple scientific representations, and apply scientific principles to everyday situations, making them highly suitable for assessing critical thinking. Although a recent study developed a critical thinking assessment specifically for temperature and heat (Sartono et al., 2025), the assessment mainly focused on cognitive measurement and did not comprehensively integrate authentic assessment tasks reflecting real-life scientific inquiry. Therefore, there remains a need for an authentic assessment instrument that measures students' critical thinking through meaningful contexts closely aligned with classroom learning and everyday phenomena.

Furthermore, effective assessment instruments should demonstrate robust psychometric quality before being implemented in educational practice. Recent validation studies have consistently emphasized the necessity of rigorous evidence for content validity, construct validity, reliability, and measurement consistency to ensure accurate interpretation of students' performance (Asriadi et al., 2026; Geschwind et al., 2024a; Wöhlke & Höttecke, 2022). Developing an authentic assessment instrument without comprehensive validation may reduce the credibility of assessment results and limit its usefulness for instructional decision-making.

Based on these considerations, this study aims to develop and validate an authentic assessment instrument for measuring senior high school students' critical thinking skills on the topic of temperature and heat at SMAN 2 Medan. The study contributes to physics education by providing a context-based assessment instrument that integrates authentic learning tasks with rigorous psychometric evaluation. The resulting instrument is expected to assist physics teachers in conducting more meaningful assessments of students' critical thinking while providing reliable evidence for improving instructional practices and supporting the development of twenty-first-century competencies.

Methods

This study was conducted to develop and validate an authentic assessment instrument for measuring senior high school students' critical thinking skills in physics

learning, particularly on the topic of temperature and heat. The development process involved a systematic sequence of needs analysis, instrument design, expert validation, field implementation, and empirical evaluation to ensure that the resulting instrument possessed adequate validity, reliability, and practicality for classroom use. The overall research procedure is described in the following sections.

Research Design

This study employed a research and development (R&D) approach to develop and validate an authentic assessment instrument for measuring senior high school students' critical thinking skills on the topic of temperature and heat. The instrument was developed using the ADDIE instructional design model, which consists of five sequential phases: Analysis, Design, Development, Implementation, and Evaluation. This model was selected because it provides a systematic framework for developing educational products through iterative design, expert validation, field testing, and continuous refinement.

Analysis Phase

The analysis phase aimed to identify the instructional and assessment needs related to physics learning at SMAN 2 Medan. A needs analysis was conducted through interviews with physics teachers to examine current assessment practices, identify the availability of authentic assessment instruments, and explore challenges in assessing students' critical thinking skills. Curriculum analysis was subsequently carried out by reviewing the Merdeka Curriculum learning outcomes and discussing their classroom implementation with teachers. In addition, a content analysis of the temperature and heat topic was performed to identify essential concepts, learning objectives, and critical thinking indicators that should be represented in the assessment instrument.

Design Phase

Based on the results of the analysis, an authentic assessment instrument was designed to measure students' critical thinking skills. The instrument consisted of 50 cause-and-effect multiple-choice items covering the concepts of temperature and heat. The item specifications were developed according to the critical thinking framework proposed by Ennis and aligned with the learning outcomes of the Merdeka Curriculum. A scoring rubric and answer key were prepared to ensure objective scoring procedures.

In addition to the test instrument, expert validation questionnaires were developed to evaluate the instrument's content relevance, construct appropriateness, and language clarity. A teacher response questionnaire was also prepared to assess the practicality and usability of the developed instrument during classroom implementation.

Development Phase

During the development phase, the assessment items were constructed according to the prepared test blueprint. The initial version of the instrument was then evaluated by an expert panel consisting of two physics education lecturers from Universitas Negeri Medan and one experienced senior high school physics teacher from SMAN 2 Medan. The experts reviewed each item in terms of content accuracy, alignment with critical thinking indicators, construct representation, language clarity, and suitability for senior high school students. Based on the experts' comments and recommendations, revisions were made to improve the quality of the assessment instrument before it proceeded to field implementation.

Implementation Phase

The revised instrument was administered to Grade XI students at SMAN 2 Medan who had completed instruction on the topic of temperature and heat. The field trial was conducted to evaluate the psychometric quality of the instrument and its feasibility for classroom use. During the implementation, students completed the assessment under teacher supervision. Subsequently, physics teachers completed the practicality questionnaire to evaluate the instrument in terms of clarity, ease of administration, suitability of content, completeness of assessment components, and overall usability.

Data Analysis

The collected data were analyzed to determine the psychometric quality of the developed instrument. Content validity was established based on expert judgments. Item validity was examined using item validity analysis, while internal consistency reliability was estimated using Cronbach's alpha coefficient. Furthermore, item discrimination indices, difficulty indices, distractor effectiveness, and overall test reliability were calculated to evaluate the quality of each assessment item. Descriptive statistics were used to analyze teachers' responses regarding the practicality of the instrument. The instrument was considered suitable for classroom implementation when it demonstrated satisfactory validity, reliability, item quality, and positive practicality evaluation.

Evaluation Phase

The evaluation phase integrated findings from expert validation, field testing, and teacher feedback to determine the overall quality of the developed instrument. Necessary revisions were made based on the empirical evidence obtained throughout the development process. The final product was an authentic assessment instrument that met the established criteria for validity, reliability, practicality, and feasibility for measuring students' critical thinking skills on the topic of temperature and heat in senior high school physics.

Result and Discussions

After the instrument product has been developed, a development trial is carried out twice, including the first trial in a small class and the second trial in a large class. Based on the results of the item validity test in small groups and large groups using the product moment correlation formula. Based on the examination and scoring carried out from student answers on small-scale trials and large-scale trials, the scores obtained from 22 students (small scale) and 36 students (large scale) are very diverse.

This study uses the feasibility test of expert validity and item validity. The expert validity test was carried out by 3 validators by assessing 3 domains, namely the content domain, construction domain and language domain. At the expert validity test stage, the analysis was carried out using Microsoft excel 2013. Based on the calculation of the expert validity test, it shows from 3 validators that of the 50 multiple choice questions reviewed, 48 questions were declared valid and 2 questions were declared invalid. This is seen based on the provision that the test instrument can be said to be valid if it meets the minimum value of the Aiken V index. After seeing the results of the data analysis, it was obtained that the V Aiken index ≥ 0.87 so that 48 questions could be used for the next test.

Questions that go through the expert validity process are continued for item trial validity. Product trials were carried out in 2 stages, namely small group trials and large group trials. According to Arikunto (2016), a good test question is valid and reliable, based on the results of the small group validity test with 22 students being the sample that from the table shows the number 0.537 obtained 35 questions declared valid (very high and high), 1 question declared invalid (high) will be revised and 12 questions declared invalid (very low) will be discarded. In line with the research of Ruth Mayasari and Nensi Nopitasari

(2022) questions that are declared invalid can be used in subsequent trials provided that there are improvements / revisions based on the trials that have been carried out.

The next feasibility analysis is the reliability test using the Kuder-Richardson 20 (KR-20) equation. From the results of research on small group trials, a value of 0.998 was obtained with a very high reliability category. Based on the reliability correlation coefficient value guided by (Arikunto, 2013) that the range of $0.91 \leq r \leq 1.00$ with very high reliability criteria, so it can be said that the overall reliability of the questions developed in the category is very high.

Analysis of the level of difficulty test there are 3 criteria, namely easy, medium, and difficult. From the results of the small group analysis obtained 9 questions with difficult criteria, 33 questions with moderate criteria and 6 questions with easy criteria. The test instrument developed is said to be good if the question is not too difficult and not too easy. If the question is too easy, the question cannot stimulate students' thinking and if it is too difficult, the student becomes discouraged and has no enthusiasm in working on the problem.

The results of the differentiation analysis, small group trials of students were divided into 2 groups, namely 50% of the upper group and 50% of the lower group. From the results of the analysis of the differentiation of small groups, 12 items were found to have poor differentiation, 10 items had sufficient differentiation, and 26 items had good differentiation.

The results of the analysis of the effectiveness of the examiner obtained are adjusted to the criteria guidelines for the effectiveness of the examiner according to (Wulaningtyas & Sukanti, 2016) based on these results it is found that 45 questions of very good examiner quality are obtained with four distractors (examiners) functioning, 2 questions of sufficient examiner quality with two distractors (examiners) functioning and 1 question of poor examiner quality with all distractors (examiners) not functioning. After the overall analysis, it is known that there are 36 test items that can be used in the next large group test.

The test instrument that has been tested, then the teacher is given a response questionnaire to see the responses or responses to the test instrument that has been developed. Teacher response analysis and as many as 2 times, which includes small groups and large groups. Based on the results of the analysis of the teacher response test for small groups consisting

of 5 teachers as respondents, an average of 87.6 was obtained in the very good category.

Large group trials were conducted widely on 36 research subjects. The results of the large group validity test with 36 students from class XI orbit being the sample that from the table shows the number 0.329 obtained 34 questions declared valid and 2 questions declared invalid questions discarded. This is proven if $r_{hitung} \geq r_{tabel}$ then the tested test instrument is valid.

The results of the research on the large group trial obtained a value of 0.946 with a very high reliability category. Based on the reliability correlation coefficient value guided by (Arikunto, 2016) that the range of $0.91 \leq r \leq 1.00$ with very high reliability criteria, so it can be said that the overall reliability of the questions developed in the category is very high. The results of research by Aiken et. al, quoted from Sumarna (2006), show that the effect of difficulty level plays the biggest role in the reliability coefficient. This is because it involves variations in the number of questions that can be answered correctly. The more difficult the questions in the test device, the greater the variation in scores obtained by the hemispheres. Thus, the greater the reliability of the test. Conversely, the lower the difficulty of a question, the smaller its reliability. For this reason, it should be avoided that many guesses are made by test takers and try to match the knowledge of test takers with the test material that will be tested on them.

The results of the large group analysis of 36 items obtained 3 difficult criteria questions (questions number 4, 9, and 39), 29 questions with moderate criteria and 4 questions with easy criteria. If the question is too easy, the question cannot stimulate students' thinking and if it is too difficult, the student becomes discouraged and has no enthusiasm in working on the problem. Then in the research on the test of differentiation in large groups, it was obtained that 3 items were declared poor differentiation, 23 items had sufficient differentiation, and 10 items had good differentiation.

The results of the analysis of the effectiveness of the triggers obtained are adjusted to the criteria guidelines for the effectiveness of the triggers according to (Wulaningtyas & Sukanti 2016) based on these results it is found that 35 questions of very good quality triggers are obtained with four distractors (triggers) functioning, and 1 question of sufficient quality triggers with two distractors (triggers) functioning.

The test instrument that has been tested, then the teacher is given a response questionnaire to see the

teacher's response or response to the test instrument that has been developed. Based on the results of the analysis of the teacher response test for a large group consisting of 5 teachers as respondents, an average of 93.6 was obtained in the good category. From these results it can be said that the teacher's response to the questions developed is very good and feasible.

Based on the results of small-scale tests and large-scale tests, the test instruments developed can be declared effective because they meet each of the criteria of good content and construct validity, balanced difficulty levels (not too easy or difficult), high differentiating power, effective checkers and high reliability. A good test instrument must be valid (measuring what should be measured), reliable (consistent and stable), have good differentiation (able to distinguish participants with high and low abilities), and proportional and appropriate difficulty levels (not too easy or difficult) so that it can reveal accurate and reliable data as a whole (Son, 2019).

Based on data analysis to measure students' critical thinking skills, it is known that of the 22 respondents in the small class trial, the percentage of students' ability to focus questions by 57% is included in the moderate category, the indicator of analyzing arguments by 55% is included in the moderate category, the indicator of concluding by 59% is included in the moderate category, then the indicator of making decisions by 45% is included in the moderate category. The percentage for the average of all indicators is 54% which can be stated in the moderate category. The results of the analysis of students' higher-level thinking skills in the large-scale test with 36 respondents showed that the percentage of students' abilities in the indicator of focusing questions by 47% was included in the moderate category, the indicator of analyzing arguments by 55% was included in the moderate category, the indicator of concluding by 51% was included in the moderate category, then the indicator of making decisions by 65% was included in the good category. The percentage for the average of all indicators is 55% which can be stated in the moderate category. The results of both small and large class trials found that SMA Negeri 16 Medan students had sufficient critical thinking skills with a percentage of 54.5%.

After the whole analysis, it is known that there are 34 questions of multiple-choice test items that are accepted and can be used because they have passed the feasibility test as a good test/question and 2 items are rejected because they do not meet the requirements of good multiple-choice test criteria from the aspects of validity, reliability, difficulty level, differentiation and effectiveness of the examiner.

Then 34 items were accepted and stored in the bank of multiple choice test questions on cause and effect in high school which became the final product in this study.

Conclusions

This study successfully developed an authentic assessment instrument for measuring senior high school students' critical thinking skills on the topic of temperature and heat. Through the ADDIE development model, the final instrument consisted of 34 cause-and-effect multiple-choice items that demonstrated satisfactory content validity, high reliability, appropriate item validity, balanced difficulty levels, acceptable discrimination indices, effective distractors, and positive teacher responses regarding its practicality. The instrument was also capable of assessing students' critical thinking across four indicators—focusing questions, analyzing arguments, drawing conclusions, and making decisions—with the overall results indicating a moderate level of critical thinking. Therefore, the developed instrument can be considered a valid, reliable, and practical assessment tool for evaluating students' critical thinking skills in senior high school physics learning, particularly on the topic of temperature and heat.

References

- Adri, J., & Abdullah, A. S. (2022). Critical Thinking Skills in Performance-Based Assessment: Instrument Development and Validation. *Journal of Technical Education and Training*, 14(1), 91–100. Scopus. <https://doi.org/10.30880/jtet.2022.14.01.008>
- Amir, et al. (2024). Pengembangan Instrumen Tes Higher Order Thinking Skills (HOTS) Peserta Didik Pada Mata Pelajaran Fisika. *Jurnal Fisika dan Pembelajarannya*, 6(2), 1-8,194-208.
- Arce-Trigatti, A., Kelley, J., & Haynes, A. (2022). On new ground: Assessment strategies for critical thinking skills as the learning outcome in a social problems course. *New Directions for Teaching and Learning*, 2022(169), 83–97. Scopus. <https://doi.org/10.1002/tl.20484>
- Arikunto, S. (2016). *Dasar-Dasar Evaluasi Pendidikan*. Jakarta Bumi Aksara.
- Asriadi, M., Hadi, S., Istiyono, E., Purnama, E., & Sanam, A. I. (2026). Assessing the Accuracy and Consistency of Formative Assessment Instruments in Physics: A Validity and Reliability Study. *International Journal of Educational Reform*. Scopus. <https://doi.org/10.1177/10567879261440527>
- Barevičiūtė, J., Dadelo, S., & Asakavičiūtė, V. (2023). The Skills of Critical Thinking, Creativity, and Communication as Tools for Overcoming Social Simulation in the Context of Sustainability: A Case Study of Students' Self-Assessment of the Affective Domain of Learning. *Sustainability (Switzerland)*, 15(14). Scopus. <https://doi.org/10.3390/su151410935>
- Daengneam, B., Sriwisathiyakun, K., & Sovajassatakul, T. (2025). Enhancing Thai General Practitioners' Critical Thinking Skills: An Educational Development Needs Assessment. *Educational Process: International Journal*, 19. Scopus. <https://doi.org/10.22521/edupij.2025.19.541>
- Damayanti, et al. (2017). Development of Authentic Assessment instruments for Critical Thinking skills in Global Warming with a Scientific Approach. *International Journal of Science and Applied Science: Conference Series*, 2(1), 289-299.
- Geschwind, G., Vignal, M., Caballero, M. D., & Lewandowski, H. J. (2024a). Evidence for validity and reliability of a research-based assessment instrument on measurement uncertainty. *Physical Review Physics Education Research*, 20(2). Scopus. <https://doi.org/10.1103/PhysRevPhysEducRes.20.020125>
- Geschwind, G., Vignal, M., Caballero, M. D., & Lewandowski, H. J. (2024b). Using a research-based assessment instrument to explore undergraduate students' proficiencies around measurement uncertainty in physics lab contexts. *Physical Review Physics Education Research*, 20(2). Scopus. <https://doi.org/10.1103/PhysRevPhysEducRes.20.020105>
- Husnawati, et al. (2019). Pengembangan Soal Higher Order Thinking Skill (HOTS) Fisika Kelas VIII SMP Materi Gerak Pada Benda. *Unnes Physics Education Journal*, 8(2), 133-140.
- Jiang, J.-P., Hu, J.-Y., Zhang, Y.-B., & Yin, X.-C. (2023). Fostering college students' critical thinking skills through peer assessment in the knowledge building community. *Interactive Learning Environments*, 31(10), 6480–6496.

- Scopus.
<https://doi.org/10.1080/10494820.2022.2039949>
- Kartowagiran, B., & Rohaeti, E. (2021). A critical thinking assessment model integrated with science process skills on chemistry for senior high school. *European Journal of Educational Research*, 10(1), 285–298. Scopus.
<https://doi.org/10.12973/EU-JER.10.1.285>
- Kumar, T., Soozandehfar, S. M. A., Hashemifardnia, A., & Mombeini, R. (2023). Self vs. Peer assessment activities in EFL-speaking classes: Impacts on students' self-regulated learning, critical thinking, and problem-solving skills. *Language Testing in Asia*, 13(1). Scopus.
<https://doi.org/10.1186/s40468-023-00251-3>
- Mayasari, R. & Nensi, Nopitasari. (2022). Pengembangan Instrumen Tes Berpikir Kreatif Tingkat Tinggi pada Peserta Didik SMP SEPREN: *Journal of Mathematics Education and Applied*, 3 (2), 102-113
- Pals, F. F. B., Tolboom, J. L. J., & Suhre, C. J. M. (2023). Development of a formative assessment instrument to determine students' need for corrective actions in physics: Identifying students' functional level of understanding. *Thinking Skills and Creativity*, 50. Scopus.
<https://doi.org/10.1016/j.tsc.2023.101387>
- Pradana, P. W. (2025). Assessment instrument of graph representations on sound wave topic: Development and measurement implementation. *Knowledge Management and E-Learning*, 17(2), 352–371. Scopus.
<https://doi.org/10.34105/j.kmel.2025.17.016>
- Rosidin, U., Sesunan, F., & Fitria, D. (2022). Development of Assessment Instruments to Measure Collaboration and Responsibility Skills of Students in Physics Learning on Cased Method-Based. *AMPLITUDO: Journal of Science and Technology Innovation*, 1(2), 29–35. Scopus.
<https://doi.org/10.56566/amplitudo.v1i2.7>
- Sappaile, et al. (2024). Pengembangan Metode Penilaian Otentik Untuk Mengukur Kemampuan Berpikir Kritis Mahasiswa. *Jurnal Review Pendidikan dan Pengajaran*, 7(1), 2139-2150.
- Sartono, B., Sunarno, W., Prayitno, B. A., & Indriyanti, N. Y. (2025). Development and Validation of A Critical Thinking Assessment on Temperature and Heat for Secondary Physics Education. *Educational Process: International Journal*, 16. Scopus.
<https://doi.org/10.22521/edupij.2025.16.264>
- Shaber, N., Shah, S. K., Imran, M., & Almusharraf, N. (2025). Exploring the Relationship Between Critical Thinking and Creativity in University Students: Gender Differences and the Assessment of Skills. *Education Sciences*, 15(4). Scopus.
<https://doi.org/10.3390/educsci15040464>
- Suarga. (2019). Hakikat, Tujuan Dan Fungsi Evaluasi Dalam Pengembangan Pembelajaran. *Jurnal Inspiratif Pendidikan*, 8 (2), 327-338.
- Wöhlke, C., & Höttecke, D. (2022). Teachers' professional vision in teaching physics—a validation study. *International Journal of Science Education*, 44(15), 2306–2329. Scopus.
<https://doi.org/10.1080/09500693.2022.2120372>
- Wulaningtyas, RC, & Sukanti. (2016). Analisis Butir Soal Ujian Akhir Semester Casual Mata Pelajaran Pengantar Akuntansi dan Keuangan. *Jurnal Kajian Pendidikan Akuntansi Indonesia*, 7, 1-18
- Zdravkovska-Adamova, B., Dzogovic, S. A., & Taneska, A. P. (2024). Application Of Standards-Based Assessment For Critical Thinking Skills During Language Acquisition. *Palimpsest*, 9(18), 185–196. Scopus.
<https://doi.org/10.46763/palim24918185za>

