Journal of English Educational Study (JEES)
Volume 8 Issue 2 November 2025 Page 176-184
E-ISSN:2655-0776

DEVELOPMENT OF CRITICAL THINKING TEST FOR THE ENGLISH SUMATIVE FINAL SEMESTER (SAS) AT ISLAMIC JUNIOR HIGH SCHOOL AS BANK OF ENGLISH TEST: A GOOGLE FORM-BASED TEST BANK

M. Saufi Rahman, M. Riffani, Mukarramah
STAI Rasyidiyah Khalidiyah
msaufirahman87@gmail.com, riffani.m@gmail.com, mukarramah21@gmail.com

Submitted: 2025-04-20
Accepted: 2025-07-09

Abstract: This study develops a critical thinking assessment for the Summative Final Semester (SAS) English Exam at Madrasah Tsanawiyah (MTs). The goal of study to know the quantitative characteristic of critical thinking test of Summative Final Semester (SAS) using google form. This study a Research and Development approach, test items were designed based on the curriculum and validated by experts. Trials test were conducted with 120 students to know difficulty, discriminant, validity, and reliability by statistical. The results indicated while students performed well on language proficiency tasks, they faced challenges with critical thinking questions, emphasizing the need for more comprehensive preparation in this area. Values of expert validation is more 0,78, validation items is > 0,3, reliability is 0.92, difficulty is easy, medium and hard, and discriminant is good. The digital platforms used Google Forms to create interactive assessments, not only measure language skills but also assess and develop critical thinking.

Keywords: Critical Thinking; Development tests; Bank of Test Items; Google Form

INTRODUCTION

Education sets a crucial role in shaping human resources that are prepared to face the challenges of the times. One of the skills increasingly required in the 21st century is critical thinking. Critical thinking is the ability to analyze, evaluate, and draw logical conclusions based on available information, as well as the ability to make sound decisions in complex situations (Liu et al., 2016). In the fact, critical thinking was considered an essential skill because it helps students not only understand the subject matter but also relate and apply it in real-life contexts. The implementation of critical thinking encompasses not only the ability to analyze and evaluate information but also the ability to solve problems, identify underlying assumptions, and assess the strengths and weaknesses of various arguments (Ennis, 1985; Hirai et al., 2022; Rustam & Priyanto, 2022). So, the critical thinking forms of the foundation for rational decision-making, based on evidence which is crucial in both academic and professional life. This highlights that education should not only focus on knowledge acquisition but also on the development of critical thinking skills, which serve as essential assets for students' future.

In Indonesia, the national education curriculum has incorporated critical thinking as a core competence that students should possess, especially in primary and secondary education. This is stated in the Regulation of the Minister of Education and Culture of the Republic of Indonesia (Kemendikbud, 2022), which emphasizes the importance of critical thinking skills in various subjects, including English. However, despite critical thinking being recognized as an essential skill, its implementation in English education at Islamic junior high school, especially Madrasah Tsanawiyah (MTs) in Indonesia remains limited.

Journal of English Educational Study (JEES)
Volume 8 Issue 2 November 2025 Page 176-184
E-ISSN:2655-0776

In the practice, assessments in English learning at MTs often focus more on mastering basic language skills, such as vocabulary, grammar, as well as speaking, reading, writing, and listening abilities. These assessments tend to neglect the measurement of students' critical thinking skills. This is a gap of evident from the practices used in many Islamic junior high school. Ideally, the assessment activity by measure students' mastery of basic knowledge, they rarely evaluate students' ability to think deeply, analyze, or evaluate the information or arguments presented (Floyd, 2011; Banta, 2015; Hofer et al., 2017;). The effective assessment should not only measure how much students memorize and repeat information but also how well they can analyze, evaluate, and construct arguments based on their understanding (Ennis, 1985; Ningsih et al., 2018; Abbasi & Izadpanah, 2018). However, in Islamic junior high school most of the tests used are still multiple-choice questions or short-answer questions that primarily measure technical aspects of the language without challenging students to think critically or apply their knowledge in broader contexts.

The testing that relies solely on memorization does not foster deep learning and, consequently, limits students' development of critical thinking skills. The effective assessment should cover various levels of thinking skills, from basic understanding to more complex critical thinking skills, such as analysis, evaluation, and creation (Szabó, 2008; Mardapi, 2016; Hirai et al., 2022). However, usually measure activities in Islamic junior high school only measure students' basic understanding of the material they have studied without assessing how they can analyze or evaluate that material in more applied contexts. Despite the rapid advancement of educational technology many Islamic junior high school still do not utilized technology to full potential to support assessments based on critical thinking.

While some schools especially Islamic junior high school have shifted to online learning after the COVID-19 pandemic, the use of technology for assessment remains limited to multiple-choice questions that only assess students' basic understanding (Seifert, 2020). The technology such as Google Forms has great potential to support the creation of more complex questions that challenge students to think more analytically, evaluatively, and creatively, but many teachers are still not accustomed to using this platform to design such questions. Castro (2018) argues that digital platforms such as Google Forms can assist teachers in creating more varied types of questions, such as those that ask students to evaluate arguments in a text or provide solutions to a given problem, which can help probe deeper into their critical thinking abilities. This platform also allows for real-time collection of data and quick feedback to students, making the assessment process more efficient. However, in reality many teachers still rely on traditional methods of assessment and underutilize digital features for designing questions that are focused on critical thinking (Adelia et al., 2021; Kurniawati & Lestari, 2020). This creates a gap in the use of technology for more comprehensive and critical-based assessments.

Given the gaps that exist in critical thinking-based assessments and the underutilization of technology for exams at MTs, this study aims to develop a critical thinking test for the final semester English exam at MTs, utilizing Google Forms as the assessment platform. This study is expected to

Journal of English Educational Study (JEES)
Volume 8 Issue 2 November 2025 Page 176-184
E-ISSN:2655-0776

address existing issues by providing an assessment tool that not only measures English comprehension but also evaluates students' critical thinking skills. Therefore, this study will help teachers design assessments that are more aligned with the goals of 21st-century education, while also supporting the development of critical thinking skills among students at Islamic junior high school.

METHOD

This study a Research and Development approach to develop a critical thinking test for the Summative Final Semester (SAS) English Exam at Islamic junior high school.. The test was designed based on the Indonesian Curriculum for eighth-grade students, with an emphasis on evaluating critical thinking skills such as analysis, evaluation, and problem-solving. The first phase involves the development of test items based on the curriculum and critical thinking indicators. The test items are developed to assess students' ability to analyse and evaluate the given information or arguments, as well as to offer solutions to relevant problems.

Then, the test items were developed, the instrument validation by a panel of six experts, including English teachers, lecturers and educational measurement specialists. The experts review the items for clarity, relevance, and their ability to assess critical thinking. The validity of each item is then calculated using the Aiken V formula, where a score above 0.78 is considered valid. The goal of this process is to ensure the test items are both appropriate and aligned with the intended assessment objectives. The data collected from the pilot test are analyzed using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to assess the underlying structure of the test and ensure that it measures the intended critical thinking skills. Cronbach's Alpha is also used to assess the internal reliability of the items, with a value above 0.7 considered acceptable.

A trial test was conducted two trials, are small scoop and wide scoop were the samples of 120 eighth-grade students from various MTs. The trial test was administered through Google Forms, facilitating easy access, data collection, and analysis. Based on the trial test results, revisions were made to improve the quality and effectiveness of the test items.

FINDING AND DISCUSSION

a. Finding

The findings of this study are based on the development, validation, and pilot testing of the critical thinking test for the Summative Final Semester (SAS) English Exam at Madrasah Tsanawiyah (MTs). Incorporating critical thinking into English lessons can enhance students' understanding and use of the language, enabling them to engage with texts and ideas more deeply. The indicators of critical thinking in English lessons, are Analyzing Literature and Texts (Close Reading, Identifying Literary Devices and Character Analysis), Developing Arguments and Opinions (Debates and Opinion Essays), Evaluating Sources (Research Projects and Fact vs Opinion), Engaging in Discussions (Socratic Seminars and Peer Review), Creative and Critical Writing (Creative Projects and Critical Reviews),

Journal of English Educational Study (JEES)
Volume 8 Issue 2 November 2025 Page 176-184
E-ISSN:2655-0776

Reflective Thinking (Journaling and Self-Assessment), and Problem-Solving Activities (Role-Playing and Case Studies).

The material indicators in accordance with the independent curriculum for class VIII for the even semester (two) consist of love our world, no littering and embrace yourself. The love our world indicator consists of several sub-indicators, namely starting of the presentation, stating the goal of the presentation, showing steps, and ending/closing the presentation. The no littering indicator consists of several sub- indicators, namely interrogative questions, past tense questions and sequencing past incidents. The embrace yourself indicator consists of several sub-indicators, namely expression of opinions and adverb of manners.

The results of the instrument validation by a panel of six experts, including English teachers, lecturers and educational measurement specialists. The calculating used the Aiken V value for each test items validity were excellent. Most of the developed items had Aiken V values above 0**.**78, indicated that the items were valid and suitable for measuring critical thinking skills. Some items that received lower ratings from the experts (for example, items with an Aiken V value below 0.78) were revised. Revisions were made to clarify the instructions, adjust the difficulty level, and correct any errors identified during the validation process. The aim was to improve the quality of the items to better measure the critical thinking aspects intended.

The analysis of the test items revealed insights into the validity, difficulty, reliability, and discriminatory power of the items. The results are summarized in the table below.

Tabel 1. Test Findings Summary

Test Item	Validity Score (Aiken V)	Difficulty Level	Discrimination Index	Cronbach's Alpha
Love Our World	0.82	Medium	0.72	0.92
No Littering	0.85	High	0.74	0.92
Embrace Yourself	0.80	Low	0.65	0.92
Environmental Impact	0.87	Medium	0.78	0.92
Problem Solving	0.90	High	0.80	0.92

An interesting exploration would be comparing the difficulty of the test items with their discrimination index. For example, items with medium or high difficulty should, ideally, show a higher discrimination index because they test higher-order thinking skills. The Difficulty Level of the test items varied. Items related to *No Littering* and *Problem Solving* were categorized as having high difficulty, while *Embrace Yourself* was found to be a lower difficulty item. This is indicative of the varying levels of challenge posed by different themes and the ability of students to apply their critical thinking skills. The Discrimination Index measures how well each item distinguishes between high and low-performing students. All items showed a positive discrimination index, with the *Embrace Yourself* item having the lowest index (0.65) and *Problem Solving* having the highest (0.80). A higher discrimination index indicates that the item is effective in differentiating between students with varying

Journal of English Educational Study (JEES)
Volume 8 Issue 2 November 2025 Page 176-184
E-ISSN:2655-0776

levels of ability. Finally, the Cronbach's Alpha coefficient for all items was consistently 0.92, suggesting high internal reliability and indicating that the items measured the same underlying construct (critical thinking) consistently across students.

Reliability Statistics

Cronbach's Alpha	N of Items
0,946	30

Figure 1. Reliability Statistic Result

b. Discussion

The development and implementation of a critical thinking test for the Summative Final Semester (SAS) English Exam at Madrasah Tsanawiyah (MTs) has provided valuable insights into the current state of critical thinking assessment in Indonesia. The findings from this study not only address the gaps identified in the Introduction but also connect to existing educational theories and frameworks that emphasize the importance of critical thinking in learning and assessment. In this section, we will discuss how the study's findings align with key theories and concepts presented earlier.

The traditional English language assessments in MTs have focused primarily on language proficiency skills, such as vocabulary, grammar, and basic comprehension. While these skills are important, they fail to assess students' ability to engage in higher-order cognitive tasks such as critical thinking. This issue is consistent with the concerns of Brookhart and Nitko (2008) who argued that assessments should go beyond factual recall and measure students' ability to analyze, evaluate, and create. The results of this study support this argument, as the critical thinking test developed in this study successfully shifted the focus of assessment from mere language proficiency to the evaluation of students' analytical and evaluative skills.

This study found that critical thinking items, such as those focused on environmental issues like *Love Our World* and *No Littering*, challenged students to apply their knowledge and think critically about the world around them. This aligns with Ennis (1985**)** definition of critical thinking, which emphasizes the ability to analyze and evaluate information or arguments. However, the results also highlighted a significant gap: while students excelled in tasks measuring language proficiency, they struggled with tasks requiring critical thinking. This suggests that while critical thinking is a goal of the Indonesian national curriculum, its integration into assessments is still lacking in practice. This gap between curriculum goals and classroom implementation was noted by Hirai et al. (2022) who highlighted that many teachers lack the skills and training necessary to integrate critical thinking into their teaching and assessment practices.

One of the challenges were the lack of emphasis on critical thinking in the teaching of English at Islamic junior high school. The findings from this study reveal that while the critical thinking test was valid and reliable, many students were not well-prepared to engage with higher-order thinking

questions. Students performed significantly better on questions that tested language proficiency but struggled with tasks that required them to analyze, evaluate, and solve problems. This difficulty in addressing more complex tasks confirms observation that students in traditional education systems are often not prepared to engage with assessments that require critical thinking. This issue was further supported by Adelia et al. (2021) who argued that assessments should measure not only knowledge recall but also higher-order thinking skills such as analysis, evaluation, and creation. This study found the students struggled with tasks requiring these higher-order thinking skills, suggesting that there has been insufficient preparation in the classroom to develop these competencies. It also supports Liu et al. (2016), who noted that the development of critical thinking skills in students requires more than just including critical thinking questions in assessments, it necessitates a pedagogical shift that prioritizes inquiry based learning, problem-solving, and reflective thinking in everyday teaching practices.

The issue of integrating technology into education, particularly in assessment, where it was noted that many teachers still rely on traditional methods for testing students' knowledge and critical thinking skills. The study findings suggest that Google Forms was an effective tool for administering the critical thinking test, as it allowed for real-time data collection and provided an efficient means of analyzing the results. This finding aligns with Seifert (2020), who emphasized the need for technology to enhance not only the administration of assessments but also the design of assessments that can evaluate complex cognitive processes. Using Google Forms allowed for the inclusion of more dynamic and interactive question types, such as those requiring students to evaluate arguments or propose solutions to real-world problems. According to Castro (2018) assertion that digital tools can facilitate the creation of assessments that go beyond simple knowledge recall and engage students in critical thinking. The fact that students performed better on language proficiency-based items and struggled with the critical thinking questions also highlights the need for professional development and training for teachers in using digital platforms to design assessments that challenge students' thinking skills.

However, despite the potential of using technology in assessments, the findings indicate that many teachers are still not fully leveraging digital tools to create more complex, critical thinking-based assessments. This further underscores the need for training and development opportunities for teachers to help them effectively integrate technology into their assessment practices (Seifert, 2020; Castro, 2018).

The analysis of the discrimination index revealed that some items effectively differentiated between high and low-performing students, while others did not. Items like *Problem Solving* and *Environmental Impact* demonstrated higher discrimination indices, suggesting that these items were more effective in assessing critical thinking. This is in line with Mardapi (2016) who argue that good assessments must strike a balance between difficulty and the ability to discriminate between students with different levels of ability. The lower discrimination index for some items, like *Embrace Yourself*, suggests that these items either did not challenge students enough or were too difficult for many to

complete effectively. This issue reflects the challenges of designing assessments that are both challenging and capable of distinguishing between high and low performers. Banta & Palomba (2015) also emphasize that an ideal test should neither be too easy nor too difficult, but should strike a balance that allows for a clear differentiation in performance. In this study, while some items succeeded in distinguishing between high and low performers, others did not, indicating the need for further refinement of the test items. Future iterations of the test will need to address this issue by adjusting the level of difficulty and ensuring that items are sufficiently challenging without being too overwhelming.

The findings of this study have important implications for future research and teaching practice found the gap between the curriculum's emphasis on critical thinking and its limited application in assessments is a pressing issue. To clearly of the future research should focus on refining the test items based on student feedback and further improving the use of technology in critical thinking assessments. Additionally, further studies could explore how different instructional strategies can enhance students' ability to perform well on assessments that require critical thinking skills. In teaching practice, the results of this study underscore the importance of integrating critical thinking into everyday lessons. Teachers need professional development in designing assessments that challenge students to think critically and apply their knowledge in meaningful ways. Furthermore, using technology in assessments should not only be seen as a means of efficiency but as a tool to enhance the depth and complexity of the tasks students are asked to perform.

CONCLUSION

The development of English items focusing on critical thinking skills for the final summative assessment of class VIII at MTs has yielded positive results. The development process involved creating indicators based on the curriculum, such as love our world, no littering, and embrace yourself. Items were then formulated and validated by a panel of experts, ensuring their accuracy and effectiveness. The validation results indicated that the items were suitable for use, with high validity values across different indicators.

The development and implementation of a critical thinking test for the Final Semester English Exam at MTs successfully addressed the gap in measuring critical thinking skills. The findings highlight the importance of integrating critical thinking into both teaching and assessment practices. While the test showed validity and reliability, the study also revealed challenges faced by students in engaging with critical thinking tasks. A Google Form platform was created to facilitate the administration of the test, reducing the use of paper and providing a more efficient way for teachers to assess students. The positive feedback from teachers and students in various schools indicated the effectiveness of the new testing approach. Overall, the development of these English test items has proven to be a valuable tool in assessing students' critical thinking skills and understanding of the curriculum.

The suggest that further efforts are needed to equip teachers with the necessary tools and skills to foster critical thinking and to ensure that assessments truly reflect students' ability to analyze, evaluate, and apply knowledge. By addressing these challenges, future assessments can play a crucial role in developing students' critical thinking abilities, which are essential for success in the 21st century.

REFERENCES

Abbasi, A., & Izadpanah, S. (2018). The relationship between critical thinking, its subscales and academic achievement of English language course: The predictability of educational success based on critical thinking. Academy Journal of Educational Sciences, 2(2), 91-105.

Adelia, A., Miftahurrahmah, M., Nurpathonah, N., Zaindanu, Y., & Ihsan, M. T. (2021). The role of google form as an assessment tool in elt: Critical review of the literature. ETDC: Indonesian Journal of Research and Educational Review, 1(1), 58-66.

Banta, T. W., & Palomba, C. A. (2015). Assessment essentials: Planning, implementing, and improving assessment in higher education. San Francisco: Jossey-Bass.

Brookhart, S. M., &Nitko, A. J. (2015). Educational Assessment of Students. In Text with Loose-Leaf Version -- Access Card Package.

Castro, S. (2018). Google Forms Quizzes and Substitution, Augmentation, Modification, and Redefinition (SAMR) Model Integration. Issues and Trends in Educational Technology, 6(2), 4–14.

Ennis, R. H. (1985). A Logical Basis for Measuring Critical Thinking Skills. In Educational Leadership

Floyd, C. B. (2011). Critical thinking in a second language. Higher Education Research & Development, 30(3), 289-302.

Hirai, A., Oka, H., Kato, T., & Maeda, H. (2022). Development and validation of an English test measuring EFL learners' critical thinking skills. Language Testing in Asia, 12(1), 45.

Hofer, S. I., Schumacher, R., & Rubin, H. (2017). The Test of Basic Mechanics Conceptual Understanding (bMCU): Using Rasch Analysis to Develop and Evaluate an Efficient Multiple Choice Test on Newton's Mechanics. International Journal of STEM Education, 4(1), 1–20.

Kurniawati, D., & Lestari, S. (2020). Using Google Forms for Online Listening Test: Does It Work?. English Education: Jurnal Tadris Bahasa Inggris, 13(2), 136-155.

Liu, O. L., Mao, L., Frankel, L., & Xu, J. (2016). Assessing Critical Thinking in Higher Education: The HEIghtenTM Approach and Preliminary Validity Evidence. Assessment and Evaluation in Higher Education, 41(5), 677– 694

Mardapi, D. (2016). Pengukuran Penilaian dan Evaluasi Pendidikan (1st ed.). Nuha Medika

Mendikbudristek, R. (2022). Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi RI Nomor 7 Tahun 2022 tentang Standar Isi pada Pendidikan Anak Usia Dini, Jenjang Pendidikan Dasar, dan Jenjang Pendidikan Menengah

Ningsih, D. R., Ramalis, T. R., & Purwana, U. (2018). Pengembangan Tes Keterampilan Berpikir Kritis Berdasarkan Analisis Teori Respon Butir. WaPFi (Wahana Pendidikan Fisika), 3(2), 45.

Rustam, R., & Priyanto, P. (2022). Critical thinking assessment in the teaching of writing Indonesian scientific texts in high school. Jurnal Penelitian Dan Evaluasi Pendidikan, 26(1), 12–25.

Seifert, T. (2020). Student Assessment in Online Learning : Challenges and Effective Practices During Covid-19. Proceedings of EdMedia + Innovate Learning, pp. 106-10(2001), 1–4.

Szabó, G. (2008). Applying Item Response Theory in Language Test Item Bank Building (Vol. 10). Peter Lang.