

A Web-Based Platform for Assessing Critical Reading Competence to Support Deep Learning

Subyantoro¹, Zuliyanti², Suseno³, Setiya Adi Buono⁴, Desyfa Cahya Aina⁵,
Leni Septiani⁶

¹²³ Faculty of Languages and Arts, Universitas Negeri Semarang

⁴⁵⁶ Undergraduate Program, Universitas Negeri Semarang

Correspondence: bintoro@mail.unnes.ac.id

HOW TO CITE:

Subyantoro, Zuliyanti, Suseno, Setiya
Adi Buono, Desyfa Cahya Aina, Leni
Septiani (2026). A Web-Based
Platform for Assessing Critical
Reading Competence to Support
Deep Learning. *International Journal
of Special Education*, 41(19s),
455 - 466

ABSTRACT:

Critical reading skills are essential for supporting deep learning, but assessments in schools are still often conducted manually and separate reading speed from comprehension. This study aims to develop and evaluate a web-based platform for measuring critical reading competence that integrates Effective Reading Speed (ERS), reading comprehension, text readability, and immediate feedback for students from elementary through high school. The study employed a research and development design consisting of six stages: needs analysis, planning, initial product development, expert validation, limited and large-scale pilot testing, and revision and implementation of the final product. The needs analysis involved 30 Indonesian language teachers. The limited pilot test involved 90 students, while the large-scale pilot test involved 210 students evenly distributed across grades 5, 8, and 11. The results showed that 93% of teachers experienced difficulties in conducting critical reading assessments and 87% needed an automated platform. All aspects of validation scored above 85%. In the large-scale pilot, Critical Reading Skills (KEM) improved descriptively by 35.3% in Grade 5, 41.3% in Grade 8, and 33.8% in Grade 11; comprehension ranged from 83–88%, while student satisfaction reached 4.5–4.7 on a 5-point scale. The novelty of this research lies in the integration of time measurement, comprehension, cross-grade readability calibration, and immediate feedback into a single Indonesian-language platform. The platform has the potential to support data-driven formative assessment, although the causal effectiveness and reliability of the instrument still require further testing.

Keywords: critical reading; deep learning; digital assessment; effective reading speed; reading literacy

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.

Open access publication under
the terms and conditions

the Creative Commons Attribution (CC
BY)

license (<http://creativecommons.org/licenses/by/4.0/>).

1 Introduction

Reading ability in a digital environment is no longer limited to understanding explicit information but encompasses source evaluation, argument analysis, information integration, and evidence-based decision-making (Kanniainen et al., 2021). Critical literacy practices also require students to recognize the interests, representations, and biases that shape a digital text (Gouseti et al., 2024). This need is increasingly important because the shift from print to screen media can alter how students process texts, although the extent of this influence depends on the design of the material, reader characteristics, and reading purposes (Furenes et al., 2021). A recent meta-analytic study indicates that technology can support reading outcomes when digital features are designed as pedagogical aids, rather than merely transferring printed text to a screen (Pasqualotto et al., 2025).

Reading literacy remains a concern at both the global and national levels, as the 2022 PISA results show a decline in average reading scores in many countries compared to the previous cycle (OECD, 2023a). Indonesian students' reading scores stood at 359, lower than the OECD average of 476 (OECD, 2023b). The proportion of Indonesian students who have not yet reached Level 2 also remains high, at 70% for girls and 79% for boys (OECD, 2023a). These data underscore that strengthening reading instruction requires assessments capable of providing diagnostic information that is rapid, measurable, and relevant to instructional decisions (Wolf & López, 2022).

Indonesia's education policy direction for 2025 positions deep learning as an approach that encourages students to understand, apply, and reflect on knowledge holistically (Ministry of Primary and Secondary Education, 2025). Deep learning is grounded in higher-order thinking activities, problem-solving, conceptual connections, and reflection on the learning process (Weng et al., 2023). A meta-analysis indicates that digital technology can enhance deep learning when used systematically and

accompanied by adequate pedagogical guidance (Wu, 2024). Technology-based formative assessment is a crucial component of this ecosystem because immediate feedback helps students monitor their understanding and refine their learning strategies (Estrada-Araoz et al., 2023).

Reading assessment in schools still faces challenges when reading speed and comprehension are evaluated separately or recorded manually. Recent developments in online reading assessment indicate that reading efficiency should be represented through both accuracy and response speed so that results reflect more than just the ability to complete tasks quickly (Yeatman et al., 2024). Adaptive and computerized assessments can also improve the efficiency of measurement and support the screening of reading difficulties among large numbers of participants (Walgermo et al., 2024). These findings align with the need for an instrument that combines reading time, accuracy of comprehension, and reporting of results into a single procedure that teachers can use routinely (Schöfl et al., 2022).

Screen use does not automatically lead to better comprehension. A large-scale analysis found that the frequency of digital device use for basic practice may be negatively correlated with comprehension, whereas targeted use for reading projects may yield a more positive relationship (Salmerón et al., 2023). Research on middle school students also indicates differences in processes and attitudes when reading on paper versus screens, although comprehension outcomes depend on task conditions and reader behavior (Jensen et al., 2024). Another meta-analysis found that general differences in comprehension between digital and print media are not always significant, but distractions, scrolling, and time pressure can hinder tasks requiring deep processing (Li & Yan, 2024). The design of assessment platforms must consider the interface, reading strategies, text complexity, and user age to ensure that technology does not increase cognitive load (Díaz et al., 2024).

Text difficulty level appropriateness is a key component in interpreting reading ability. The use of readability measures helps ensure alignment between text demands and the reader's developmental level, although readability formulas cannot replace assessments of content, culture, and learning objectives (Fry, 1968). Online assessments that align reader ability and text complexity on the same scales can provide more meaningful diagnostic information (Orellana et al., 2024). A good reading instrument must also explicitly test constructs, include indicators that represent various comprehension processes, and provide evidence of validity for the intended use of the scores (Trimiño-Pérez et al., 2024). Comparisons of test formats indicate that results on tablets and paper can be reasonably comparable, but test order, language, and participant characteristics still need to be controlled for (Seifert & Paleczek, 2022).

Previous research has yielded technology-based reading assessments, adaptive screening tools, and formative feedback systems (Walgermo et al., 2024). Technology-based academic assessments have provided evidence of validity for formative use among language learners (Wolf & López, 2022). Online instruments for reading efficiency have been developed with a strong focus on psychometric accuracy and scalability (Yeatman et al., 2024). Digital formative assessments have also been shown to support motivation, self-regulation, and reading engagement in specific contexts (Estrada-Araoz et al., 2023). However, these studies generally focus on a single language, a single grade level, a single reading component, or a single learning context; thus, they have not yet simultaneously integrated reading skills, comprehension, the readability of Indonesian-language texts, thematic variety, immediate feedback, and cross-grade-level use from elementary through high school (Trimiño-Pérez et al., 2024).

This empirical gap highlights the need for a local platform that connects pedagogical design,

automated assessment, and cross-grade-level diagnostic data (Yeatman et al., 2024). Methodological gaps are also evident in the limited number of development studies that report on teacher needs, expert validation, phased pilot testing, changes in critical reading skills, comprehension, satisfaction, and user experience within a single evaluation sequence (Wolf & López, 2022). This study aims to develop a web-based platform for measuring critical reading competencies in Indonesian and to evaluate its needs, validity, practicality, and initial usage outcomes (Walgermo et al., 2024). The study's contribution lies in an assessment model that integrates reading time, comprehension accuracy, cross-grade readability, theme classification, and immediate feedback to support deep learning (Wu, 2024). Based on these gaps and objectives, the research questions were formulated to address the need to align constructs, tasks, and score interpretations (Trimiño-Pérez et al., 2024): (1) What needs underpin the platform's development, and what are its design characteristics? (2) how do users perceive the platform's expert validity and practicality; and (3) what are the descriptive changes in KEM, comprehension levels, and student satisfaction in grades 5, 8, and 11?

2. Methodology

Research Design

The study employed a research and development approach adapted from the Borg and Gall model (Gall et al., 2003). The model was simplified into six stages to align with the development of a web-based educational assessment tool: needs analysis; design planning; initial product development; expert validation; limited and large-scale pilot testing; and revision and implementation of the final product. Development was conducted iteratively, as feedback from validators and users was used to refine the text, test items, user interface, and reporting functions.

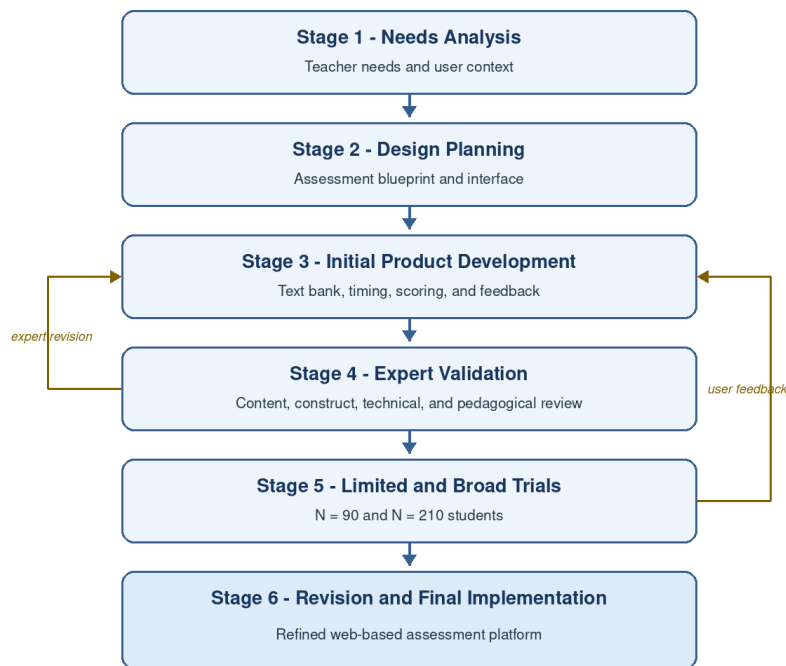


Figure 1. Research and Development Flow

Figure 1 illustrates the relationships among the development stages, data sources, and outputs for each stage. Needs analysis yields user specifications; the design and development stages yield a prototype; validation and testing yield evidence of feasibility; and final revisions yield a platform ready for use in formative assessment.

Population, Sample, and Research Context

The research subjects consisted of three groups. The first group comprised validators with expertise in literacy, Indonesian language education, educational technology, and learning assessment. The second group consisted of Indonesian language teachers as the platform's primary users. The third group comprised fifth-grade elementary school students, eighth-grade junior high school students, and eleventh-grade senior high school students from several schools in Central Java Province.

Participants were selected using purposive sampling, ensuring representation across grade levels and stages of reading-ability development. The needs analysis involved 30 Indonesian language teachers. A limited pilot test involved 90 students, 30 students from each grade level, to assess the clarity of instructions, system

functionality, and initial challenges. A large-scale pilot test involved 210 students, 70 students each from grades 5, 8, and 11, to obtain an overview of KEM results, understanding, and user satisfaction.

Product and Measurement Framework

The developed product is a web platform that includes a reading text bank, automatic time tracking, multiple-choice comprehension questions, KEM calculation, and immediate feedback. Texts were classified based on the Dewey Decimal Classification system and calibrated for difficulty using the Fry Scale. The selection of themes was intended to broaden the variety of discourse, while readability calibration served as an initial check of the texts' appropriateness for each grade level.

KEM is calculated by combining reading speed and comprehension percentage. The operational formula used is $KEM = (\text{number of words} / \text{reading time in seconds} \times 60) \times (\text{comprehension score} / 100)$. This score is displayed automatically after students complete the questions, allowing teachers to obtain data on reading efficiency, rather than just the number of words per minute.

Data Collection Techniques and Instruments

Quantitative data were obtained from teacher needs questionnaires, expert validation sheets, reading time records, comprehension test scores, KEM results, and a five-point Likert-scale satisfaction questionnaire. Qualitative data were obtained through observations of user interactions, field notes, teacher interviews, and students' open-ended responses. A combination of these data was used to evaluate the product across content, technical aspects, pedagogy, and user experience.

The research instruments consisted of: (1) a teacher needs analysis questionnaire; (2) an expert validation sheet; (3) a collection of reading texts calibrated using the Fry Scale; (4) multiple-choice comprehension questions; (5) a student satisfaction questionnaire; and (6) observation and interview guidelines. The content validity of the texts and test items was assessed through expert judgment. The validation sheets covered reading content, item construction, technical feasibility, and pedagogical appropriateness. A score of more than 85% was defined as "highly valid" according to the product evaluation criteria used in this study.

Internal reliability coefficients for the test and questionnaire were not available in the source data. This study therefore makes no claims regarding reliability based on Cronbach's alpha, omega, or other psychometric indices. This limitation is noted as the basis for the recommendation that future studies report internal consistency, score stability, item analysis, and inter-rater reliability when open-ended responses are used.

Research Procedures

The first phase mapped teachers' challenges, feature requirements, and user characteristics. The second phase translated these needs into information architecture, interface design, database specifications, text categories, and mechanisms for time tracking and assessment. The third phase produced a prototype capable of reading text, recording duration, answering

questions, calculating KEM, and displaying results.

The fourth stage involves asking validators to evaluate the content, structure, technical aspects, and pedagogical suitability. Revisions are made based on the validators' comments before a limited pilot test. The fifth stage involves testing the prototype with 90 students, addressing usability issues, and then conducting a large-scale pilot test with 210 students. The sixth stage involves refining the platform and formulating recommendations for its implementation in formative assessment.

Data Analysis Techniques

Data on needs, validation, reading proficiency, comprehension, and satisfaction were analyzed descriptively using percentages and means. Changes in reading proficiency were calculated using the formula $((\text{post-test score} - \text{pre-test score}) / \text{pre-test score}) \times 100\%$. The analysis was differentiated by grade level so that patterns in grades 5, 8, and 11 could be compared without losing the developmental characteristics of the readers.

Data from observations, interviews, and open-ended responses were analyzed thematically through repeated reading, initial coding, code grouping, theme review, and the compilation of a narrative of findings. The main themes focused on ease of use, text variety and relevance, feedback, motivation, and diagnostic benefits. The available data did not include standard deviations, inferential effect sizes, or individual data; therefore, the pre-test–post-test improvement was interpreted as a descriptive change rather than evidence of causal effectiveness.

3. Results

Teacher Needs and Product Specifications

The needs analysis revealed a strong demand for more efficient and integrated reading assessments. A total of 93% of 30 teachers reported difficulty in effectively conducting critical reading assessments due to time

constraints, inaccuracies in manual record-keeping, and the lack of instruments aligned with curriculum needs. A total of 87% of teachers wanted a web-based platform that automatically calculates KEM and provides immediate results for formative evaluation.

Interviews reinforced the questionnaire results. Teachers need texts with diverse themes, difficulty level classifications, individual result reports, and actionable information to determine follow-up actions. These specifications were translated into a tiered text bank, automatic time tracking, comprehension questions, KEM calculations, and post-test feedback.

Expert Validation

Expert evaluation covers the appropriateness of reading content, item construction, technical feasibility, and pedagogical suitability. All aspects received an average score above 85% and were categorized as highly valid based on research criteria. Validators determined that the text was aligned with students' cognitive development, the items focused on analytical and evaluative

processes, the interface was understandable, and the feedback mechanisms were relevant for formative assessment.

The validation process identified several areas for revision, including simplifying instructions, ensuring consistency in terminology, adjusting readability levels, re-examining distractors, and improving the clarity of the results display. These revisions were implemented prior to the large-scale pilot test. The validation results support the suitability of the content and initial design, but do not replace the need for psychometric testing based on student response data.

Descriptive Changes in Effective Reading Speed

The large-scale pilot test involved 210 students divided equally across three grade levels. Students read texts with adjusted difficulty levels and then answered comprehension questions. Reading times were automatically recorded, and ERR scores were calculated by the system. A summary of pre-test and post-test scores is presented in Table 1.

Table 1. Comparison of Pre-Test and Post-Test Effective Reading Speed

Education Level				Pre-Test KEM (wpm)	Post-Test KEM (wpm)	Improvement (%)
Elementary Grade)	School	(5th	85.2		115.3	35.3
Junior High Grade)	School	(8th	132.7		187.5	41.3
High School (Grade 11)				218.4	292.4	33.8

Table 1 shows a descriptive increase across all grade levels. The largest relative increase occurred in Grade VIII at 41.3%, followed by Grade V at 35.3% and Grade XI at 33.8%. The simple average across groups of equal size changed from 145.4 to 198.4 words per minute, representing an increase of approximately 36.4%. These figures are not interpreted as statistically significant differences because individual data and measures of variation are not available.

Reading Comprehension and User Satisfaction

Comprehension and satisfaction were examined to ensure that the increase in speed was not accompanied by a decrease in answer accuracy and to assess user acceptance. Table 2 presents the comprehension and satisfaction scores for each grade level.

Table 2. Reading Comprehension and User Satisfaction

Education Level	Average Reading Comprehension (%)	Student Satisfaction Score (1–5)
Elementary School (5th Grade)	83	4.5
Junior High School (Grade 8)	85	4.6
High School (Grade 11)	88	4.7

Table 2 shows that comprehension was above 80% at all grade levels. Scores increased gradually from 83% in fifth grade to 88% in eleventh grade. Student satisfaction ranged from 4.5 to 4.7 on a 5-point scale. This pattern indicates high user acceptance during the pilot phase, though it does not imply that satisfaction is synonymous with learning effectiveness.

Qualitative Findings

Observations, open-ended responses, and interviews yielded four main themes. These themes summarize students' experiences and teachers' perspectives during the use of the platform and are presented in Table 3.

Table 3. Qualitative Themes from User Experience

Theme	Evidence from Users	Observed Product Value
Interface clarity	Students completed the sequence with minimal assistance	Supports independent test completion
Text relevance and variety	Students described the texts as familiar and interesting	Encourages engagement across themes
Immediate feedback	Students appreciated seeing their results immediately after completing the test	Supports self-monitoring and correction
Diagnostic value	Teachers identified differences in individual reading performance	Supports follow-up instruction and remediation

The findings in Table 3 show that the simple interface, contextual texts, and immediate results were the most frequently appreciated features. Teachers saw the platform's potential for mapping individual needs and periodic monitoring. Development feedback primarily focused on the need for a class dashboard, progress history, expansion of the text bank, and more detailed reporting.

4. Discussion

Platform Needs and Design Characteristics

The first research question examined the platform's development needs and design characteristics. The percentage of teachers reporting assessment difficulties indicates a gap between reading learning demands and the

capacity of manual procedures. This finding aligns with the development of technology-based classroom assessments that prioritize administrative efficiency and formative utility as the foundation for score interpretation (Wolf & López, 2022). Automated timing systems are also relevant to advancements in reading efficiency measurement that combine speed and accuracy so that results are not reduced to mere mechanical fluency (Yeatman et al., 2024).

The platform's design integrates a text bank, readability, thematic categories, timing, comprehension questions, KEM, and immediate feedback. This integration distinguishes the product from apps that merely provide text and quizzes or tools that measure only fluency. The use of technology that is integrated with pedagogical goals has greater potential to support deep learning than the use of fragmented tools (Wu, 2024). Findings by Salmerón et al. (2023) also remind us that the frequency of tool use does not in itself improve comprehension; therefore, the quality of tasks and the function of feedback must be central to the design.

Calibrating text using the Fry Scale provides an initial procedure for adjusting text difficulty to grade level. The use of this formula must be accompanied by content evaluation, as sentence length and syllable count do not fully account for prior knowledge, discourse structure, culture, and reading objectives. An approach that aligns text difficulty with reader ability can strengthen the interpretation of results and prevent students from being assessed using texts that are too easy or too difficult (Orellana et al., 2024). Further development should combine readability, item analysis, and empirical performance data, as done in the validation of primary reading proficiency tests (Trimiño-Pérez et al., 2024).

Validity and Practicality of the Platform

The second research question addresses expert validity and practicality. Validation scores above 85% across all aspects provide preliminary evidence that the content, construction, technical aspects, and pedagogical design were deemed acceptable by experts. These results align with the principle that digital assessments must be built on

the basis of arguments regarding score usage, construct validity, and the learning context, not merely the functionality of the application (Wolf & López, 2022). The development of adaptive screening tests also highlights the importance of the connection between task design, psychometric evidence, and teachers' decision-making needs (Walgermo et al., 2024).

Satisfaction scores of 4.5–4.7 indicate high practicality and acceptance. Students rated the interface as easy to use, the text as varied, and the immediate results as helpful in identifying errors. Computerized formative assessments can enhance engagement, autonomy, and self-regulation when feedback is understandable and can guide subsequent actions (Estrada-Araoz et al., 2023). Alazemi's (2024) findings also suggest that technology-integrated formative assessment has the potential to support understanding and the learning experience, although results from a foreign language context cannot be directly generalized to Indonesian language learning.

The practicality of the platform is further supported by findings that tablet and paper-based modes can yield fairly comparable scores when test conditions are controlled (Seifert & Paleczek, 2022). Comparisons of media must still account for test order, reading habits, language characteristics, and screen size. Research by Jensen et al. (2024) indicates that students may exhibit different attitudes and processing patterns on screens versus paper, although differences in results are not always consistent. School implementations need to provide a brief orientation, adequate devices, stable connections, and accessibility procedures to ensure that scores are not compromised by technical issues.

Expert validity alone is insufficient to establish that an instrument possesses well-established psychometric qualities. Research by Yeatman et al. (2024) demonstrates that online instruments must be tested for reliability, convergent validity, measurement precision, and the ability to track progress. Research by Schöfl et al. (2022) also emphasizes predictive validity and empirical feasibility before screening tools are used for high-stakes decisions. The platform in this study

is more appropriately positioned as a low-stakes formative tool until internal consistency, item functioning, cross-grade invariance, and score stability have been tested.

Changes in KEM, Understanding, and Satisfaction Across Grade Levels

The third research question addressed descriptive changes in reading efficiency, comprehension, and satisfaction. Improvements in reading efficiency occurred across all grade levels, with the largest relative change observed in 8th grade. This pattern may indicate that integrating time tracking, structured practice, and feedback helps students regulate their reading pace. These findings are consistent with the notion that reading efficiency should be measured through the ability to accurately comprehend a text within a reasonable time limit (Yeatman et al., 2024). The results also align with evidence that digital interventions can benefit reading skills when their features, intensity, and target readers are appropriately designed (Pasqualotto et al., 2025).

Comprehension scores in the 83–88% range indicate that the increase in speed during the trial was not accompanied by a decrease in average comprehension. This finding is important because fast reading without meaning construction does not meet the goals of critical literacy. Research by Habók et al. (2024) shows that online reading strategies are associated with comprehension outcomes and require explicit instruction. Future platforms could incorporate metacognitive prompts, evidence highlighting, reflective questions, and answer explanations so that assessment does not stop at the final result.

Differences in improvement across grade levels should be interpreted with caution. Research by Furenes et al. (2021) indicates that age and digital design characteristics moderate reading outcomes on screens. A review by Díaz et al. (2024) emphasizes that the outcomes of technology use are influenced by pedagogical approaches and technology design, not the medium alone. The higher improvement observed in 8th grade may be related to initial growth potential, device usage habits, text appropriateness, or responses to the

test format; the available data cannot yet distinguish among these explanations.

Findings across grade levels also provide preliminary support for the platform's contribution to deep learning. Critical reading provides the foundation for activities involving understanding, connecting, evaluating, and reflecting on information. Weng et al. (2023) explain that deep learning requires authentic tasks that promote problem-solving and higher-order thinking. This platform currently measures only some of these prerequisites, namely, efficiency and text comprehension; the claim that the platform directly fosters deep learning requires additional measurements of argument analysis, source evaluation, knowledge transfer, and reflection.

Theoretical and Practical Implications

Theoretically, this study demonstrates that KEM can be positioned as a formative assessment construct that links fluency and comprehension, further enriched by readability and user experience. This model expands digital assessment from mere administrative automation toward a learning information system. The critical digital literacy framework emphasizes that technical skills must be accompanied by the ability to assess the meaning, sources, and significance present in a text (Gouseti et al., 2024). Future item banks need to increase the proportion of tasks involving evidence analysis, source comparison, bias detection, and credibility evaluation so that the term “critical reading” is more strongly represented.

In practice, teachers can use the platform for initial assessment, periodic monitoring, grouping by needs, and determining follow-up actions. The classroom dashboard should display trends in reading speed and comprehension separately so that teachers can distinguish between students who read slowly but comprehend well and those who read quickly but have difficulty understanding. The findings of Walgermo et al. (2024) support the use of adaptive tests to improve the precision and efficiency of screening, while Wolf and López (2022) emphasize that formative reports must be

actionable for teachers. Integration with learning tools should still allow for print reading, discussions, teacher explanations, and strategy practice.

Research Limitations

This study has several limitations. The sample was selected purposefully and included only schools in Central Java; therefore, national generalizations cannot yet be made. The available data do not include detailed demographic characteristics, network quality, device types, standard deviations, individual data, or a control group. The descriptive pre-test–post-test design cannot distinguish the platform’s influence from practice, maturity, text differences, or other learning factors.

The reliability of the tests and questionnaires has not been reported, while validity is primarily based on expert judgment. The Fry Scale only estimates surface readability and does not guarantee equivalence in conceptual or cultural complexity. Critical reading assessments are also still dominated by multiple-choice questions, so the processes of source evaluation, argumentation, and reflection have not been measured in depth. These limitations require that the results be understood as evidence of initial feasibility and practicality, not as confirmation of causal effectiveness.

5. Conclusion

The study produced a web-based platform for measuring critical reading competency that integrates time tracking, KEM, comprehension,

cross-grade readability calibration, thematic variation, and immediate feedback. Needs analysis indicates high demand from teachers; expert validation falls into the “highly valid” category; and extensive pilot testing showed a descriptive increase in KEM of 33.8–41.3%, with comprehension at 83–88% and satisfaction ratings of 4.5–4.7. The scientific contribution and novelty of this research lie in the integration of measurement components, which were previously often developed separately, into a single local platform for 5th-, 8th-, and 11th-grade students. These results support the platform’s feasibility as a low-risk formative assessment tool for helping teachers identify patterns in student progress.

Future research should employ a quasi-experimental or experimental design with a comparison group, a multisite sample, and individual-level data to test the platform’s effects using inferential methods. Subsequent studies should report on internal consistency, item analysis, convergent validity, measurement invariance, and score stability, while also developing a teacher dashboard, improving accessibility, and incorporating tasks for evaluating sources and arguments. Integrating data on reading processes, reading strategies, and open-ended critical tasks will strengthen the platform’s position, transforming it from a measure of reading efficiency into a comprehensive critical reading assessment system that supports deep learning.

6. References

- Alazemi, A. F. T. (2024). Formative assessment in artificial integrated instruction: Delving into the effects on reading comprehension progress, online academic enjoyment, personal best goals, and academic mindfulness. *Language Testing in Asia*, 14, Article 44. <https://doi.org/10.1186/s40468-024-00319-8>
- Díaz, B., Nussbaum, M., Greiff, S., & Santana, M. (2024). The role of technology in reading literacy: Is Sweden going back or moving forward by returning to paper-based reading? *Computers & Education*, 213, 105014. <https://doi.org/10.1016/j.compedu.2024.105014>
- Estrada-Araoz, E. G., Sayed, B. T., Niyazova, G. G., & Lami, D. (2023). Comparing the effects of computerized formative assessment vs. computerized dynamic assessment on developing EFL learners’ reading motivation, reading self-concept, autonomy, and self-regulation. *Language Testing in Asia*, 13, Article 39. <https://doi.org/10.1186/s40468-023-00253-1>
- Fry, E. (1968). A readability formula that saves time. *Journal of Reading*, 11(7), 513–516, 575–578.

- Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research*, 91(4), 483–517. <https://doi.org/10.3102/0034654321998074>
- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction* (7th ed.). Allyn & Bacon.
- Gouseti, A., Lakkala, M., Raffaghelli, J., Ranieri, M., Roffi, A., & Ilomäki, L. (2024). Exploring teachers' perceptions of critical digital literacies and how these are manifested in their teaching practices. *Educational Review*, 76(7), 1751–1785. <https://doi.org/10.1080/00131911.2022.2159933>
- Habók, A., Oo, T. Z., & Magyar, A. (2024). The effect of reading strategy use on online reading comprehension. *Heliyon*, 10(2), e24281. <https://doi.org/10.1016/j.heliyon.2024.e24281>
- Jensen, R. E., Roe, A., & Blikstad-Balas, M. (2024). The smell of paper or the shine of a screen? Students' reading comprehension, text processing, and attitudes when reading on paper and screen. *Computers & Education*, 219, 105107. <https://doi.org/10.1016/j.compedu.2024.105107>
- Kannianen, L., Kiili, C., Tolvanen, A., Aro, M., Anmarkrud, Ø., & Leppänen, P. H. T. (2021). Assessing reading and online research comprehension: Do difficulties in attention and executive function matter? *Learning and Individual Differences*, 87, 101985. <https://doi.org/10.1016/j.lindif.2021.101985>
- Ministry of Primary and Secondary Education. (2025). *Ministry of Primary and Secondary Education Regulation No. 13 of 2025 reinforces policy direction through deep learning*. <https://gtk.kemendikdasmen.go.id/news/siaran-pers/c034d26f-83b6-4b4b-926c-b572e665e477>
- Li, Y., & Yan, L. (2024). Which reading comprehension is better? A meta-analysis of the effect of paper versus digital reading in the last 20 years. *Telematics and Informatics Reports*, 14, 100142. <https://doi.org/10.1016/j.teler.2024.100142>
- OECD. (2023a). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- OECD. (2023b). *PISA 2022 results: Country note—Indonesia*. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/indonesia_c2e1ae0e-en.html
- Orellana, P., Silva, M., & Iglesias, V. (2024). Students' reading comprehension level and reading demands in teacher education programs: The elephant in the room? *Frontiers in Psychology*, 15, 1324055. <https://doi.org/10.3389/fpsyg.2024.1324055>
- Pasqualotto, A., Cunningham, E. G., Holman, C., Bediou, B., & Bavelier, D. (2025). Digital tools for reading success: Meta-analyses of digital interventions. *Journal of Cognitive Enhancement*, 9(4), 455–479. <https://doi.org/10.1007/s41465-025-00336-2>
- Salmerón, L., Vargas, C., Delgado, P., & Baron, N. (2023). The relationship between the use of digital tools in the language arts classroom and reading comprehension scores. *Reading and Writing*, 36(1), 175–194. <https://doi.org/10.1007/s11145-022-10295-1>
- Schöfl, M., Steinmair, G., Holzinger, D., & Weber, C. (2022). Predicting word reading deficits using an app-based screening tool at school entry. *Frontiers in Pediatrics*, 10, 863477. <https://doi.org/10.3389/fped.2022.863477>
- Seifert, S., & Paleczek, L. (2022). Comparing tablet and print modes of a German reading comprehension test in third grade: Influence of test order, gender, and language. *International Journal of Educational Research*, 113, 101948. <https://doi.org/10.1016/j.ijer.2022.101948>
- Trimiño-Pérez, L., Hurtado-Reina, J., Velasco, E., & Martínez, A. (2024). Design and validation of a Primary Reading Proficiency Test (PCL-P). *Frontiers in Language Sciences*, 3, 1471040. <https://doi.org/10.3389/flang.2024.1471040>
- Walgermo, B. R., Foldnes, N., Uppstad, P. H., Bakken, A. M., & Lundetræ, K. (2024). Development and deployment of an adaptive national elementary reading screening test. *Frontiers in Education*, 9, 1331777. <https://doi.org/10.3389/feduc.2024.1331777>
- Weng, C., Chen, C., & Ai, X. (2023). A pedagogical study on promoting students' deep learning through design-based learning. *International Journal of Technology and Design Education*, 33, 1653–1674. <https://doi.org/10.1007/s10798-022-09789-4>
- Wolf, M. K., & López, A. A. (2022). Developing a technology-based classroom assessment of academic reading skills for English language learners and teachers: Validity evidence for formative use. *Languages*, 7(2), 71. <https://doi.org/10.3390/languages7020071>
- Wu, X.-Y. (2024). Exploring the effects of digital technology on deep learning: A meta-analysis. *Education and Information Technologies*, 29, 425–458. <https://doi.org/10.1007/s10639-023-12307-1>
- Yeatman, J. D., Tran, J. E., Burkhardt, A. K., Ma, W. A., Mitchell, J. L., Yablonski, M., Gijbels, L., Townley-Flores, C., & Richie-Halford, A. (2024). Development and validation of a rapid and precise online sentence reading efficiency assessment. *Frontiers in Education*, 9, 1494431. <https://doi.org/10.3389/feduc.2024.1494431>

Example:

7. Acknowledgements (Optional)

- Mention any funding, assistance, or institutional support

8. Author Declaration

- Confirm originality of work
- State no conflicts of interest