



A Teacher-Made Test for an Opinion Paragraph: Insights from Design-Based Research (DBR)

Mukminatus Zuhriyah^{1*}, Nur Mukminatien², Soo Kum Yoke³, Suhono⁴

¹ Universitas Hasyim Asy'ari Tebuireng Jombang, Indonesia

² Universitas Negeri Malang, Indonesia

³ Universiti Teknologi MARA Cawangan Negeri Sembilan, Malaysia

⁴ Universitas Ma'arif Lampung, Indonesia

zoehrea@gmail.com*

Abstract

In English as a Foreign Language (EFL) teaching and learning, teachers frequently design their own tests to measure students' skills, monitor progress, and evaluate instructional effectiveness. However, most previous research on teacher-made tests in EFL writing has centered on essay types such as descriptive, argumentative, narrative, expository, and comparison-contrast, leaving opinion paragraphs relatively underexplored. Addressing this gap, the present study describes the systematic development and validation of a teacher-made test specifically designed to assess opinion paragraphs. Conducted in a paragraph writing course at a private university in East Java, the study adopts a genre-based approach and draws on the design-based research (DBR) framework proposed by Cavallaro and Sembiante (2021). The instrument consists of two components: a writing prompt and an analytic scoring rubric. Its quality was ensured through a comprehensive validation process, including content and face validity (involving two writing instructors and eight students) and inter-rater reliability testing by two trained raters. The resulting test provides a cohesive, valid, and reliable formative assessment tool for evaluating EFL learners' opinion paragraph writing. The inter-rater reliability coefficient ($r = 0.85$) indicates a strong positive correlation, supporting the consistency and reliability of the scoring rubric. This study not only contributes to the enhancement of teachers' assessment literacy but also offers a practical model for developing and validating classroom-based writing assessments in EFL contexts.

Keywords: A Teacher-Made Test, Design-Based Research, Formative Assessment, Genre-Based Approach

ARTICLE INFO

Article history:

Received
August 03, 2025
Revised
September 19,
2025
Accepted
October 28, 2025

Published by

Website

ISSN

Copyright



CV. Creative Tugu Pena

<https://attractivejournal.com/index.php/anglophile/>

2746-8631

This is an open access article under the CC BY SA license

<https://creativecommons.org/licenses/by-sa/4.0/>

@ 2025 by the authors

INTRODUCTION

The pursuit of quality education, particularly in the realm of English as a foreign language (EFL) writing, is inextricably linked to the implementation of effective assessment practices. It is evident that, as time has passed, the paradigm of assessment has undergone substantial evolution (Khotimah et al., 2024). Initially dominated by a summative orientation, often termed Assessment of Learning (AoL), which primarily focused on measuring student performance through objective evidence like scores, the field has progressed to embrace initiatives that prioritize the use of assessment to actively promote student learning, known as Assessment for Learning (AfL) and a more advanced approach, Assessment as Learning (AaL), encouraging students to engage in self and peer

assessment, fostering their metacognitive skills and self-regulation. Related to AfL, Fitriyah et al. (2022) state that the development of effective learning is a responsibility shared by teachers and students and AfL supports teachers in delivering effective teaching and learning. Meanwhile, AaL, a subtype of AfL, primarily relies on students' inclinations and capacities to cultivate metacognition to establish objectives, track and assess past learning using internal feedback, and produce new learning (Schellekens et al., 2021).

A substantial body of research in EFL writing classrooms posits that when EFL writing classroom assessment shifts towards AfL and AaL, students' writing skills improve considerably, aligning with teachers' expectations. For instance, several studies report that AfL is significant to improve students' writing skills since it gives students the ability to take control of writing assessments (Alshakhi, 2019; Fitriyah et al., 2022; Zou et al., 2021). Other studies on the implementation of AaL reveal that students gain better writing quality and efficiency, more evaluation and feedback literacy, and a stronger sense of resilience and ownership as writers as a result of the AaL-oriented training (He & Wang, 2024; Wang & Lee, 2021; Xiang et al., 2022). This standpoint emphasizes the pivotal role of effective classroom assessment in the effective teaching of EFL writing, underscoring the necessity for a more pronounced emphasis on teachers' assessment literacy in the development of customized assessment instruments.

Glušac and Milić (2022) in their systematic literature review report that the instructions of teacher-made tests were deemed sufficiently clear and concise, and they were primarily written in English as the foreign language of study. Besides, this review study also explicitly states that teacher-made tests provide information directly relevant to the test's purpose, procedures, and task aspects often lacking in standardized tests. However, these instruments frequently exhibit inherent weaknesses, often demonstrating a lack of sophisticated design knowledge and skill on the part of teachers and are commonly not subjected to adequate validation processes (Majeed & Jasim, 2022; Pesiwarissa et al., 2021; Tewachew et al., 2024). This observation highlights a pressing need to bolster teachers' assessment literacy, equipping them with the expertise to construct valid and reliable self-made assessment tools. The research often looks at teacher-made assessment instruments, focusing on their validity and reliability (Gashaye & Degwale, 2019; Puspitasari & Ulyasari, 2023; Rohmah, 2019) as well as different kinds of evidence, namely content, cognitive, and authentic evidence (Pesiwarissa et al., 2021). Nevertheless, they lack a comprehensive, step-by-step explanation of the development and validation process. Related to the research on teacher-made tests focusing on the text types were also still limited to a descriptive text (Santi et al., 2019), an argumentative text (Mukminatien et al., 2021), a narrative and expository text (Jeong, 2017), and a comparison and contrast essay (Khotimah et al., 2024). This gap is particularly pronounced for specific text types, such as an opinion paragraph, especially when approached through a genre-based pedagogical framework.

Therefore, the present study aims to address this empirical void by meticulously detailing the development and quality maintenance of a teacher-made test, specifically designed for an opinion paragraph. By providing an explicit account of this process, the study seeks to offer practical contributions to the advancement of teachers' assessment literacy, enabling them to self-design and develop effective assessments for their writing instruction. To achieve the overarching objectives, the study is guided by two principal research questions:

1. How is an opinion paragraph test developed?
2. How is an opinion paragraph test' quality maintained?

LITERATURE REVIEW

1. Roles, Principles, and Challenges of a Teacher-Made Test

Teachers play a central and multifaceted role in the development and implementation of assessments, including in teacher-made tests to ensure that such assessments align with instructional goals and respond to students' learning needs (Yen et al., 2023). Their responsibilities go beyond merely creating test items; teachers must also determine the test's purpose, select appropriate content and formats, and design tasks that reflect real classroom practices and objectives. In this capacity, teachers function as curriculum interpreters, instructional designers, and assessors, making decisions that directly influence the validity, reliability, and authenticity of classroom assessments (DeLuca et al., 2016). Furthermore, teacher-made tests provide a flexible platform for formative assessment, enabling educators to adjust teaching strategies based on student performance data (Misra, 2021). By tailoring test design to reflect both content knowledge and students' language proficiency, especially in EFL contexts, teachers contribute significantly to promoting meaningful learning and fair evaluation. As such, the construction of teacher-made tests becomes not only an evaluative activity but also a reflective professional practice that enhances assessment literacy and pedagogical effectiveness.

To fulfill this complex role effectively, teachers must also be guided by foundational assessment principles that ensure the quality and integrity of the tests they develop. These principles consisting of validity, reliability, practicality, and authenticity serve as essential criteria for evaluating the effectiveness of teacher-made assessments (Brown & Abeywickrama, 2019). It can be said that the first principle is validity comprising of content validity, face validity, construct validity, and criterion-related validity, which ensure alignment with intended learning outcomes, while reliability as the second principle emphasizing consistent and dependable results. Then, the third is practicality referring to the feasibility of test implementation within typical classroom constraints, such as time, resources, and grading workload. Authenticity, especially crucial in EFL writing tasks, as the final principle, promotes real-world relevance by assessing students in contexts that mirror actual communicative use. Integrating these principles allows teachers to design tests that are pedagogically sound and contextually appropriate, ultimately enhancing the impact of assessment on student learning (Pesiwarissa et al., 2021). By combining these principles with a reflective, learner-centered approach, teachers are empowered to create classroom assessments that are both rigorous and responsive to the diverse needs of their students.

While adherence to key assessment principles can enhance the quality of teacher-made tests, teachers often face significant challenges in putting these principles into practice. One of the most frequently reported difficulties is a lack of formal training in assessment design, which can lead to misunderstandings about how to ensure validity, balance item difficulty, or create reliable scoring criteria (Majeed & Jasim, 2022; Setiabudi et al., 2019; Tewachew et al., 2024). In many EFL contexts, teachers may also struggle with developing authentic and contextually relevant tasks, particularly when test formats are influenced by outdated or standardized models that prioritize accuracy over communicative effectiveness (Glušac & Milić, 2022). Additionally, constraints such as limited time, insufficient resources, and high workload can hinder teachers from engaging in reflective and collaborative test development (Majeed & Jasim, 2022; Tewachew et al., 2024). These challenges are further compounded when teachers are expected to design assessments without institutional support or access to professional development opportunities. As a result, despite their central role in the assessment process, teachers may rely on intuitive or inherited practices, which can limit the pedagogical effectiveness of the instruments they create. Recognizing and addressing these challenges is therefore

essential to empowering teachers to construct high-quality assessments that are aligned with learning goals and responsive to student needs.

2. An Opinion Paragraph in a Genre-Based Approach

An opinion paragraph is a short, structured piece of writing in which the writer expresses a personal viewpoint on a debatable topic and supports it with logical reasoning, examples, or explanations. It typically includes a topic sentence that clearly states the writer's stance, supporting sentences that provide reasons and evidence, and a concluding sentence that restates the opinion or summarizes the main point. This paragraph type plays a critical role in developing students' argumentative thinking and persuasive communication skills, especially in English as a Foreign Language (EFL) context. It serves as a foundational genre that prepares learners for more extended forms of academic writing, such as opinion essays, argumentative essays, and reflective papers. By mastering the opinion paragraph, students build not only writing fluency but also confidence in expressing personal viewpoints in a structured and coherent way (Hyland, 2019).

Within the genre-based approach, an opinion paragraph is treated as a socially situated text with a specific communicative purpose, audience, and structure. This approach emphasizes that writing is not merely a set of grammar rules or sentence patterns but a way of engaging in purposeful communication shaped by context. In genre-based pedagogy, students are explicitly taught the generic features of an opinion paragraph, its organization, language patterns, and rhetorical moves, through a scaffolded process that includes modeling, joint construction, and independent writing. The teaching of opinion paragraphs, therefore, involves raising students' awareness of how ideas are organized, how language is used persuasively, and how to adjust their writing for purpose and audience. Research has shown that this explicit, genre-focused instruction can significantly enhance EFL students' ability to structure arguments and use cohesive devices appropriately (Nagao, 2022; Zhai & Razali, 2023).

Integrating the opinion paragraph into a genre-based approach carries important pedagogical implications. First, it encourages genre awareness, helping students recognize not only how to write an opinion paragraph, but also why certain structures and language choices are expected in particular contexts. This awareness supports transferable writing skills, allowing learners to adapt their rhetorical strategies to different genres across academic and professional domains. Second, a genre-based approach promotes equity in the classroom by making the expectations of academic writing transparent and accessible to all learners, including those from linguistically diverse backgrounds. For teachers, this approach provides a clear instructional framework for planning writing lessons and designing assessments that reflect authentic language use. Ultimately, the genre-based teaching of opinion paragraphs contributes to more meaningful, student-centered instruction by linking form, function, and context in a way that fosters both language development and critical thinking (Hyland, 2019; Tardy, 2019).

METHOD

Since the primary objectives of this study were to examine the development and quality assurance of a classroom assessment instrument for writing an opinion paragraph for first-year EFL undergraduate students at a private university in East Java, Indonesia, design-based research (DBR) inspired by Cavallaro and Sembianti (2021) was employed. DBR was chosen for its suitability in comprehensively studying instructional aspects while simultaneously designing and refining an instructional product, allowing for iterative development and refinement based on practical application. The DBR framework typically encompasses four main procedures: 1) pre-implementation; 2) design; 3) implementation and revision; and 4) reflection and evaluation.

The application of the DBR phases to the development of the opinion paragraph assessment instrument proceeds as follows. Phase 1 is pre-implementation. This initial phase involves a thorough examination of the existing paragraph writing course description and syllabus. The objective is to gain familiarity with how opinion paragraph writing is contextualized and taught within the curriculum. This detailed analysis serves as the foundational step for constructing the specific assessment goal, identifying relevant indicators, and developing a comprehensive blueprint for the instrument. Then, phase 2 is design. In this phase, the core assessment instrument for opinion paragraph writing is developed. This includes crafting a writing prompt designed to elicit an opinion paragraph and analytical scoring rubric tailored to assess their specific features. Initial feedback on the writing prompt is systematically gathered from fellow writing teachers to ensure its clarity and relevance. While phase 3 is implementation and revision. The newly developed instrument is, then, administered to a separate group of students in a pilot study. During this tryout, students' comments and feedback regarding the writing prompt and the overall test administration process are carefully collected. Subsequently, two raters are recruited and undergo specific training to ensure consistent application of the scoring rubric when rating the students' opinion paragraphs. Following the rating process, inter-rater reliability is estimated, and necessary revisions to the instrument are made based on all gathered feedback and reliability analysis. Phase 4 is reflection and evaluation. The final phase involves a comprehensive evaluation and reflection on the entire process of instrument development and the resulting product. This integrates findings from all preceding phases to draw conclusions about the instrument's quality and the development process itself. The details of how this research design is integrated into assessment instrument development can be seen in Figure 1.

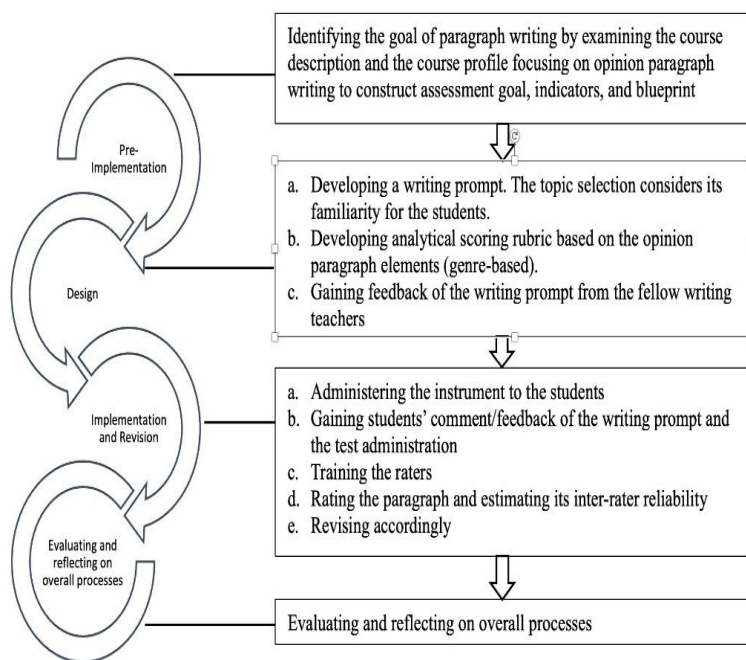


Figure 1. DBR phases inspired by Cavallaro and Sembiant (2021) in the assessment instrument development

The participants involved in this study and their respective roles were carefully selected to ensure the robustness of the development and validation process. There were 2 fellow teachers, 8 students, and 2 trained raters. The two fellow writing teachers participated in this study, providing written comments on the developed assessment instrument, which was an opinion paragraph test. Their feedback focused on the clarity of

language in instructions, the relevance of the content to instructional objectives and indicators, and the practicality of the format, including time allotment and organization. While the eight students were a group of eight undergraduate students, from a different cohort but possessing a similar beginning to intermediate level of language proficiency as the target students, who were involved in the implementation phase. They completed an opinion paragraph writing task to express their opinion whether the obligation for students at schools to wear uniforms is mostly beneficial or mostly restrictive, using the developed prompt and subsequently provided feedback through a survey and short reflection on aspects such as readability, time adequacy, and test relevance. Then, the two trained raters were recruited to blind-rate the student-produced opinion paragraphs independently. They underwent thorough training to ensure consistent and reliable application of scoring rubric. Both the two fellow teachers and two raters were experienced writing teachers, each with over five years of teaching experience in writing, particularly paragraph writing courses. The selection of all participants was based on their direct relevance to the study's objectives and their accessibility within the academic context. All the participants in this study have confirmed their agreement to be involved in this study in their consent forms provided.

Data collection procedures encompassed various sources to provide a comprehensive understanding of the instrument's development and quality. They consisted of self-observational data (the authors, acting as the assessment instrument developers, maintained self-observational records throughout all four DBR phases, documenting decisions, challenges, and refinements), eight written opinion paragraphs produced by the students during the pilot study, written feedback from the two fellow writing teachers and the eight students regarding the assessment instrument's various aspects, and the scores assigned by the two trained raters to the students' opinion paragraphs. All the data were collected, analyzed, and interpreted to answer the research questions.

For data analysis, a mixed-methods approach was employed. In qualitative analysis, common patterns and themes from the written feedback provided by both fellow teachers and students were identified. This involved careful transcription, repeated reading, systematic coding, and interpretive analysis. Simple calculations from brief surveys were also performed to discern inclusive patterns. While in quantitative analysis, the Pearson correlation coefficient formula was utilized to estimate the inter-rater reliability. This statistical analysis determined the consistency of the two sets of scores generated by the independent raters. The correlation result was interpreted against a standard table for correlation size. In the final reflection and evaluation phase, the findings from both qualitative and quantitative analyses were integrated to provide a holistic interpretation of the overall results. The iterative nature of DBR, with its continuous cycle of design, implementation, revision, and reflection, serves as a robust quality assurance mechanism. This continuous feedback loop, involving multiple stakeholders such as fellow teachers, students, and raters, ensures that the assessment instrument is not a static product but undergoes ongoing improvement based on practical application and expert and user input. This iterative refinement process directly contributes to enhancing the instrument's validity and reliability over time.

RESULT AND DISCUSSION

An opinion paragraph development in this study reported the hypothetical result (0.85). It demonstrated a high degree of positive correlation for inter-rater reliability. It means that the opinion paragraph test made in this study is valid and reliable that can be used as a reference for other English writing teachers in producing teachers-made writing tests. The detailed explanation of the development of this opinion paragraph test can be presented as follows.

1. The Development of an Opinion Paragraph Test

The initial phase of developing the writing prompt and scoring rubric for an opinion paragraph writing commenced with a meticulous identification of the course description for the paragraph writing course. This process was crucial for establishing the content validity of the instrument, ensuring it aligned with the broader instructional goals of the curriculum. The course description is as follows.

This course develops students' skills in writing clear and well-structured paragraphs across various types, including narrative, descriptive, expository, and opinion. It covers essential paragraph elements such as topic sentences, supporting details, and concluding sentences, as well as the principles of unity and coherence.

The course description indicated that the paragraph writing course aimed to develop students' abilities in various paragraph types, with opinion paragraph writing being a specific component. This deliberate analysis served as the foundational step for formulating the assessment goal and its corresponding indicators.

a. The goal of an opinion paragraph test

The assessment aims to evaluate students' ability to write a well-organized opinion paragraph by clearly stating a position on a debatable topic (whether the obligation for students at schools to wear uniforms is mostly beneficial or mostly restrictive) and supporting it with logical reasoning and relevant examples, using appropriate paragraph structure and language.

b. Indicators of an opinion paragraph test

Based on the assessment goal and a comprehensive review of literature on opinion paragraph writing, the following key indicators were established to evaluate students' writing:

1. Apply the generic structure of an opinion paragraph by including: (a) a clear topic sentence stating the opinion, (b) supporting sentences with logical reasons and relevant examples, and (c) a concluding sentence that restates the opinion or summarizes the main idea.
2. Use appropriate language, including effective vocabulary (e.g., persuasive language, linking words) and accurate grammatical structures (e.g., modal verbs, cohesive devices).
3. Construct sentences and transitions effectively to create logical relationships between ideas and maintain coherence.
4. Demonstrate control of mechanics, including correct spelling, punctuation, and capitalization.
5. Provide sufficient and specific support for the opinion, using logical reasoning, examples, or explanations appropriate to the writing level.

In opinion paragraph writing, content assessment goes beyond topic relevance. It involves evaluating the strength of argumentation, the logical flow of ideas, and the clarity and credibility of the support provided for the writer's stated opinion.

c. Development of the Analytical Scoring Rubric

The scoring rubric was meticulously developed, directly reflecting the defined assessment goal and indicators, and explicitly integrating the principles of a genre-based approach to opinion paragraph writing. The criteria within the rubric were designed to encompass both the distinctive generic structure of opinion paragraphs and common writing conventions. The rating scale ranged from 1 to 100, with each performance indicator assigned a specific weight to reflect its relative importance in the overall score. A minimum score of 25 was set to acknowledge students' effort in completing the writing task.

Assigning different weights to rubric indicators is a deliberate pedagogical decision that influences both what students prioritize in their learning and what raters emphasize during assessment. In opinion paragraph writing, where the strength of

argument and the quality of support are central, giving greater weight to the “Body Section: Argument and Evidence” signals its instructional significance. This decision shapes the teaching-learning process by encouraging students to develop persuasive reasoning and guiding raters to focus on evaluating argumentation as a core skill.

The blueprint of the opinion paragraph assessment instrument is detailed in Table 1, providing a transparent and comprehensive breakdown of the assessment criteria, their alignment with instructional objectives, and their relative importance. This table serves as a foundational document for test developers, raters, and students, ensuring consistency and objectivity in scoring and clarifying expectations.

Table 1. Blueprint of an opinion paragraph test

Assessment goal: The assessment aims to evaluate students’ ability to write a well-organized opinion paragraph by clearly stating a position on a debatable topic and supporting it with logical reasoning and relevant examples, using appropriate paragraph structure and language.		
Paragraph Component (The components were formulated based on the instructional objectives represented in the assessment goal)	Criteria (indicators) for a qualified opinion paragraph	Weight in %
Topic Sentence and Introduction	A clear topic sentence expressing a direct opinion; includes the topic and engages the reader.	15
Supporting Points and Reasoning	Logical, relevant points that support the opinion. May include brief examples or explanations. Focus on clarity and connection to the opinion.	35
Concluding Sentence	Restates opinion in new words, summarizes key point(s), ends with a thoughtful wrap-up.	15
Language Use (Vocabulary and Grammar)	Correct grammar, appropriate and varied vocabulary, clear expression of ideas.	15
Organization and Transitions	Clear progression of ideas; use of transition words/phrases to connect points and enhance flow.	15
Mechanics	Correct spelling, punctuation, and capitalization.	5

d. Design of an opinion paragraph writing prompt

The writing prompt was designed based on selected relevant and debatable topics, ensuring they were familiar and closely related to classroom materials to enable students to effectively express their opinions. The prompt was structured to provide clear instructions and expectations for the opinion paragraph. Figure 2 shows the opinion paragraph writing prompt.

An Opinion Paragraph Writing Prompt

(45 Minutes)

Direction

1. In many schools, students are required to wear uniforms. Some educators argue that this rule promotes equality, discipline, and school identity. However, others claim that the obligation to wear uniforms is restrictive and may limit students' individuality and self-expression. Do you think the obligation for students to wear uniforms is mostly beneficial or mostly restrictive?
 - a. Write one paragraph (150–200 words) expressing your opinion.
 - b. Start with a topic sentence stating your opinion.
 - c. Include at least two reasons to support your opinion.
 - d. Use examples or explanations to support your points.
 - e. Finish with a sentence that summarizes your main idea.
2. Your paragraph will be evaluated by the following aspects: clear structure of the opinion paragraph, language use (vocabulary and grammar), transitions and sentence variety, mechanics (spelling, punctuation, capitalization), and how well you support your opinion.

You have 45 minutes to write your opinion paragraph without looking up the dictionary.

Figure 2. An Opinion Paragraph Writing Prompt

e. Administering the Instrument (an opinion paragraph test)

In this phase, two fellow writing teachers gave feedback which was instrumental in making necessary revisions to enhance its clarity and effectiveness. Following these revisions, a pilot study was conducted with eight undergraduate students. This try-out aimed to evaluate the instrument's appropriateness, clarity, and the adequacy of the allocated time. The administration process was carefully observed to identify any practical issues. After completing the writing task, students were asked to complete a brief survey and write a short reflection on their experience, providing valuable qualitative data on their perceptions of the test.

f. Rating the paragraphs and estimating inter-rater reliability

After the pilot study, the eight collected opinion paragraphs were blind rated independently by two trained raters. Prior to the actual rating, the raters underwent comprehensive training. This training involved a detailed review of each performance indicator within the scoring rubric, a demonstration of how to assess a sample paragraph, and practice scoring sessions to ensure a consistent understanding and application of the rubric. Once consistency was established, they proceeded to score the students' paragraphs. To determine the consistency of the two sets of scores produced by the raters, the inter-rater reliability was estimated using a Pearson correlation coefficient. The result of this calculation shows a high positive correlation between the two sets of scores.

2. Maintaining the Quality of the Opinion Paragraph Test

Maintaining the quality of the assessment instrument (an opinion paragraph test) was a central focus throughout its development, primarily achieved through rigorous validation processes. In this case, the test developers employed content and face validity and inter-rater reliability. The content validity of the developed teacher-made test for opinion paragraph writing was meticulously ensured by confirming its direct congruence with the established course objectives and the institution's paragraph writing course syllabus. Test developers conducted a thorough review of existing course catalogs and syllabi to verify that the instrument accurately measured the intended opinion paragraph writing skills, knowledge, and abilities of the students.

While assessing face validity, two fellow writing teachers were asked to provide their opinions on the assessment instrument sheet, focusing on its language clarity (whether instructions were clear and unambiguous), content relevance (whether the test

content aligned with instructional objectives and assessment goals), and format (whether time allotment was sufficient and organization was complete). Overall, they provided positive feedback, predominantly ticking 'Yes' for all elements, indicating general satisfaction with the instrument. They also offered constructive notes for further improvement, particularly suggesting making the indicators more explicit to guide students better in their writing. The two writing fellow teachers' feedback for the developed writing assessment instrument (an opinion paragraph test) are compiled in Table 2.

Table 2. Fellow writing teachers' feedback

Fellow Writing Teachers (FWT)	Good points highlighted (+)	Points needing improvement (-)
FWT 1	This opinion paragraph test is well-designed. I think its strengths lie in the clarity of language, the relevance of its content to an opinion paragraph, and its appropriate format.	The writing prompt needs to refer more explicitly to the indicators in detail, especially regarding the use of language and logical connectors for student' opinions. I suggest also including a brief explanation of the expected depth and specificity of examples required to support students' opinion. This would help students understand the level of detail necessary.
FWT 2	I think the test instrument is well-structured and quite complete for assessing an opinion paragraph. The prompt is engaging.	My suggestion is that the test instrument will become better if the expected word count for an opinion paragraph (e.g., 150-200 words) is clearly stated, and students are reminded about the accuracy of their opinion.

The two writing teachers generally agreed that the assessment tool was well-developed in terms of language, content, and format. They agreed on a suggestion to improve the quality of the assessment instrument by being more precise about its indicators. We improved our writing assessment instrument after receiving input from fellow teachers. During the piloting phase, we discovered that the suggested evaluation time of 45 minutes was sufficient for students to finish their writing tasks.

The next step was conducting a pilot study with eight students to evaluate readability, time allotment, and exam relevancy to the materials learnt. The pilot study was conducted to gather their perceptions on the instrument's readability, the adequacy of time allotment, and its relevance to the learned material. All participating students unanimously agreed that the assessment instrument was clear and unambiguous, that the time provided was sufficient, and that the test content was relevant to the materials they had learned. Students' reflections indicated that while some initially felt nervous about being assessed, their comfort improved significantly after understanding that the assessment's primary aim was formative, without immediate academic consequences. The students' opinions about the assessment instrument (an opinion paragraph test) are shown in Table 3.

Table 3. Students' participants' opinions about the opinion paragraph test

Statements	Students' responses		
	Yes	No	Not Sure
Language (The instruction and writing prompt are clear and free from ambiguity for an opinion paragraph.)	100%	0%	0%

Content (The test content is relevant to the materials learned on an opinion paragraph.)	100%	0%	0%
Format and test time (The format is appropriate and practical to follow. The test time allotment is reasonably sufficient.)	100%	0%	0%

The students also expressed their ideas on the opinion paragraph test. Most of them were anxious about being tested, but after getting an explanation about the purpose of the test, which had no academic repercussions, they felt considerably better.

After the pilot study, the eight collected opinion paragraphs were scored by two trained raters. It was conducted to see the consistency of the scoring (reliability of the test). A Pearson correlation coefficient analysis was performed on the two sets of scores obtained from Rater 1 and Rater 2. The results, presented in Table 4, indicate a high degree of positive correlation between the two sets of scores, demonstrating strong consistency in their application of the developed scoring rubric.

Table 4. Pearson correlation coefficient result

		Rater 1	Rater 2
Rater 1	Pearson Correlation	1	.85**
	Rater 1	Pearson Correlation	1
	N	8	8
Rater 2	Pearson Correlation	.85**	
	Sig. (2-tailed)	.001	1
	N	8	8

** Correlation is significant at the 0.01 level (2-tailed).

The correlation coefficient of 0.85 ($p = 0.001$) signifies a very high positive correlation, indicating a substantial level of agreement between the trained raters in assessing and judging the same opinion paragraphs based on the same developed scoring rubric. While a high inter-rater reliability confirms scoring consistency, a deeper examination of component-level scores and any subtle discrepancies between raters can yield valuable diagnostic information about student learning. For example, if students consistently score lower on the "strength of evidence" or "logical flow" criteria in their opinion paragraphs, or if raters show minor disagreements on these specific aspects, it signals areas where instructional reinforcement might be particularly beneficial. This means the reliability analysis extends beyond a mere validation check; it functions as a continuous feedback loop for pedagogical improvement, allowing educators to identify specific student learning gaps and tailor targeted interventions.

DISCUSSION

The findings of this Design-Based Research (DBR) demonstrate how a teacher-made test for an opinion paragraph was systematically developed to adhere to the characteristics of a high-quality classroom assessment tool, specifically practicality, reliability, validity, and authenticity, as advocated by (Brown & Abeywickrama, 2019; Suskie, 2018). The iterative nature of the DBR process, moving through design, implementation, revision, and reflection, and critically involving multiple stakeholders, fellow writing teachers, students, and trained raters, played a pivotal role in enhancing the instrument's overall quality (McKenney & Reeves, 2018). The continuous feedback loop from these participants significantly contributed to the instrument's refinement and robustness (Wang et al., 2023).

The instrument's design, deeply rooted in a genre-based approach, inherently ensured its authenticity. By emphasizing the social context, purpose, and structure of an opinion paragraph, the assessment directly measures students' ability to produce paragraphs that are meaningful and relevant to real-world communicative situations (Hyland, 2019; Tardy, 2019). The development process successfully incorporated specific

components crucial to assessing the opinion paragraph, such as the quality of argumentation and evidence, and the effective use of appropriate transitional signals, aligning with recommendations from relevant literature on opinion writing (Insuwan & Thongrin, 2025; Tran, 2021). The high inter-rater reliability, with a correlation coefficient of 0.85, further confirms the instrument's readiness for practical application in classroom settings.

This project underscores the feasibility for teachers to design high-quality assessment instruments, thereby demonstrating a mastery of assessment literacy. A crucial aspect highlighted is the indispensable coherence between the genre-based teaching approach and the assessment principles employed. This alignment is fundamental for fostering optimal student learning in opinion writing (Hyland, 2019; Pastore, 2023). The detailed explanation of this instrument's development serves as a valuable reference, directly addressing common challenges faced by teachers in constructing rigorous assessment instruments, such as a perceived lack of design knowledge or an over-reliance on traditional, less effective methods (Brown et al., 2024; Matsumoto-Royo & Ramírez-Montoya, 2021).

The process of designing a detailed, genre-specific assessment instrument with explicit criteria, such as those developed for an opinion paragraph test, compels educators to critically examine their curriculum and instructional practices. If the assessment consistently reveals student weaknesses in areas critical to an opinion paragraph writing, such as difficulty crafting a clear topic sentence, providing sufficient and relevant support, or using persuasive language, it serves as a powerful signal for a re-evaluation of teaching content, methods, or focus (Hyland, 2019; Suskie, 2018). The assessment thus acts as a diagnostic tool not only for individual students but also for the overall teaching-learning system, prompting a review of instructional coherence and effectiveness (Fan et al., 2021; Harris et al., 2022). For instance, if student performance data indicated consistent struggles with providing strong supporting reasons or using logical structure, it would suggest the need for writing teachers to reinforce these skills in daily instruction, using focused practice on paragraph-level opinion writing strategies.

Despite its strengths, this study has some limitations. First, the validation process involved a small number of participants. It only used two writing teachers and eight students that may limit the generalizability of the results. Second, there were only two raters who tested the inter-rater reliability. It may not fully represent the consistency that could be achieved with a larger group. Third, because the test was developed and validated in a particular educational context, its effectiveness might differ in other settings with different curricula or student backgrounds. These limitations suggest that further studies with larger and more diverse samples are needed to strengthen the reliability and applicability of the instrument.

To further advance this area, it is recommended that future research empirically investigates the effectiveness of employing this developed instrument in authentic classroom settings through relevant research designs, such as case studies or experimental research. Additionally, conducting similar design-based research studies focusing on the process of developing assessment instruments for other language skills or specific writing genres (e.g., argumentative essays, narrative paragraphs) would further enrich the existing literature on teachers' assessment literacy development. A well-structured process for developing assessment instruments, coupled with robust validation and reliability procedures, will undoubtedly enhance teaching practices in writing courses and EFL instruction more broadly.

,

CONCLUSION

This article has shed light on the careful and deliberate development of a teacher-made test for an opinion paragraph, meticulously designed within a genre-based approach. The project produced a comprehensive set of formative assessment tools, including a writing prompt and detailed scoring rubric. Through a systematic design-based research process, incorporating observations of assessment administration, feedback from fellow writing teachers and students, and rigorous estimation of inter-rater reliability, the developed instrument was confirmed to be practical, reliable, valid, and authentic. The statistical analysis, employing a Pearson correlation coefficient, yielded a coefficient of 0.85 ($p = 0.001$), signifying a very high positive correlation and demonstrating a substantial degree of agreement between the trained raters in their assessment of the same opinion paragraphs based on the same developed scoring rubric. The development of this assessment instrument in practice offers a meaningful contribution to improving teachers' assessment literacy, especially in the context of teaching an opinion paragraph writing to EFL undergraduate students. It also stands as a robust model for other educators who seek to develop similar assessments and for future researchers aiming to expand the scholarship on classroom assessment.

REFERENCES

- Alshakhi, A. (2019). Revisiting the writing assessment process at a Saudi English language institute: Problems and solutions. *English Language Teaching*, 12(1), 176–185. <https://doi.org/10.5539/elt.v12n1p176>
- Brown, H. Do., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (Third Edit). Pearson Education, Inc.
- Brown, T. D., Barnes, M., & Finefter-Rosenbluh, I. (2024). Teacher perspectives and experiences of assessment literacy in Victorian junior secondary schools. *Australian Journal of Education*, 68(1), 5–22. <https://doi.org/10.1177/00049441231214022>
- Cavallaro, C. J., & Sembiante, S. F. (2021). Facilitating culturally sustaining, functional literacy practices in a middle school ESOL reading program: A design-based research study. *Language and Education*, 35(2), 160–179. <https://doi.org/10.1080/09500782.2020.1775244>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248–266. <https://doi.org/10.1080/10627197.2016.1236677>
- Fan, T., Song, J., & Guan, Z. (2021). Integrating diagnostic assessment into curriculum: A theoretical framework and teaching practices. *Language Testing in Asia*, 11(1), 1–23. <https://doi.org/10.1186/s40468-020-00117-y>
- Fitriyah, I., Bastomi, Y., Khotimah, K., & Gozali, I. (2022). Implementation of assessment for learning in online EFL writing class: A case of novice undergraduate teachers. *LEARN Journal: Language Education and Acquisition Research Network*, 15(2), 129–159.
- Gashaye, S., & Degwale, Y. (2019). The content validity of high school English language teacher made tests: The case of Debre work preparatory school, East Gojjam, Ethiopia. *International Journal of Research in Engineering, IT and Social Sciences*, 9(11), 41–50.
- Glušac, T., & Milić, M. (2022). Quality of written instructions in teacher-made tests of English as a foreign language. *English Teaching and Learning*, 46(1), 39–57. <https://doi.org/10.1007/s42321-021-00079-1>
- Harris, L. R., Adie, L., & Wyatt-Smith, C. (2022). Learning progression-based assessments: A systematic review of student and teacher uses. *Review of Educational Research*, 92(6), 996–1040. <https://doi.org/10.3102/00346543221081552>
- He, M., & Wang, L. (2024). Implementing assessment as learning in online EFL writing

- classes. *RELC Journal*, 1–15. <https://doi.org/10.1177/00336882231224966>
- Hyland, K. (2019). *Second language writing* (Second Edi). Cambridge University Press.
- Insuwan, C., & Thongrin, S. (2025). Empowering Thai EFL learners as critical thinkers and skilled writers : A genre-based approach with critical pedagogy. *REFlections*, 32(1), 487–520. <https://doi.org/10.61508/refl.v32i1.280405>
- Jeong, H. (2017). Narrative and expository genre effects on students, raters, and performance criteria. *Assessing Writing*, 31, 113–125. <https://doi.org/10.1016/j.asw.2016.08.006>
- Khotimah, K., Batunan, D. A., & Mukminatien, N. (2024). Teachers' made writing assessment instrument: Pedagogical insight from process to product. *Studies in English Language and Education*, 11(1), 95–115. <https://doi.org/10.24815/siele.v11i1.26774>
- Majeed, N. T., & Jasim, J. K. (2022). An evaluation of the English teacher-made tests for the fifth preparatory class of Iraqi schools. *Journal of Tikrit University for Humanities*, 29(3, 2), 1–28. <https://doi.org/10.25130/jtuh.29.3.2.2022.22>
- Matsumoto-Royo, K., & Ramírez-Montoya, M. S. (2021). Core practices in practice-based teacher education: A systematic literature review of its teaching and assessment process. *Studies in Educational Evaluation*, 70(September), 1–13. <https://doi.org/10.1016/j.stueduc.2021.101047>
- McKenney, S., & Reeves, T. (2018). *Conducting educational design research* (Second Edi). Routledge.
- Misra, P. K. (2021). Learning and teaching for teachers. In *Learning and Teaching for Teachers*. Springer Nature Singapore. <https://doi.org/10.1007/978-981-16-3077-4>
- Mukminatien, N., Yaniafari, R. P., El Khoiri, N., & Ariyanto, M. S. A. (2021). The rater reliability of an argumentative writing assessment using genre-based scoring rubric. In N. Hayati, Suharyadi, S. Andreani, & U. P. Astuti (Eds.), *The changing face of ELT: A festschrift for Prof. Ali Saukah and Prof. M. Adnan Latief* (Vol. 11, Issue 1, pp. 175–191). Universitas Negeri Malang. http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEmbetungan_Terpusat_Strategi_Melestarl
- Nagao, A. (2022). A Genre-based Approach to teaching descriptive report Writing to Japanese EFL university students. *Tesl-Ej*, 26(3), 1–28. <https://doi.org/10.55593/ej.26103a13>
- Pastore, S. (2023). Teacher assessment literacy: a systematic review. *Frontiers in Education*, 8(July), 1–25. <https://doi.org/10.3389/feduc.2023.1217167>
- Pesiwarissa, R., Anaktototy, K., & Maruanaya, H. J. (2021). An analysis of EFL teacher-made tests Content knowledge, cognitive, and authentic evidences. *MATAI: International Journal of Language Education*, 2(1), 1–13. <https://doi.org/10.30598/matail.v2i1.5485>
- Puspitasari, N. L., & Ulyasari, N. (2023). An analysis of validity and reliability of a teacher-made test on descriptive text: A case study to tenth grade office administration students at SMKN 3 Pontianak. *Jurnal Sinestesia*, 13(2), 1297–1303.
- Rohmah, N. (2019). Validity and reliability study on teacher-made assessment for English mid-term examination. *Eleventh Conference on Applied Linguistics (CONAPLIN 2018)*, 254(Conaplin 2018), 107–110. <https://doi.org/10.2991/conaplin-18.2019.236>
- Santi, E., Haris Setiawan, M., & Khumaedi, M. (2019). Developing assessment instrument on writing on descriptive paragraph based on picture of junior high school students. *Journal of Educational Research and Evaluation*, 8(2), 116–122. <https://doi.org/10.15294/jere.v8i2.36086>
- Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van

- der Vleuten, C. P. M. (2021). A scoping review on the notions of assessment as learning (AaL), assessment for learning (AfL), and assessment of learning (AoL). *Studies in Educational Evaluation*, 71(October), 1–15. <https://doi.org/10.1016/j.stueduc.2021.101094>
- Setiabudi, A., Mulyadi, & Puspita, H. (2019). An Analysis of Validity and Reliability of A Teacher-Made Test. *Journal of English Education and Teaching (JEET)*, 3(4), 522–532.
- Suskie, L. (2018). *Assessing student learning: A common sense guide* (Third Edit). Jossey-Bass.
- Tardy, C. (2019). *Genre-based writing: What every ESL teacher needs to know*. University of Michigan Press ELT.
- Tewachew, A., Shiferie, K., & Tefera, E. (2024). Practices of EFL teachers in test construction. *Cogent Education*, 11(1), 1–18. <https://doi.org/10.1080/2331186X.2024.2412496>
- Tran, M. T. T. (2021). Use of self-regulated learning strategies in paragraph writing at Van Lang university. *International Journal of TESOL & Education*, 1(3), 1–13. <http://eoi.citefactor.org/10.11250/ijte.01.03.001>
- Wang, L., & Lee, I. (2021). L2 learners' agentic engagement in an assessment as learning-focused writing classroom. *Assessing Writing*, 50(October), 1–16. <https://doi.org/10.1016/j.asw.2021.100571>
- Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023). Chinese EFL teachers' writing assessment feedback literacy: A scale development and validation study. *Assessing Writing*, 56(April), 1–16. <https://doi.org/10.1016/j.asw.2023.100726>
- Xiang, X., Yuan, R., & Yu, B. (2022). Implementing assessment as learning in the L2 writing classroom: A Chinese case. *Assessment and Evaluation in Higher Education*, 47(5), 727–741. <https://doi.org/10.1080/02602938.2021.1965539>
- Yen, P. H., Thuy, P. T., Thao, L. T., Thi, N. A., Tra, N. H., & Thu, H. T. A. (2023). Multi-faceted identities of teacher assessors: A qualitative study in the Vietnamese educational context. *Asian Journal of Social Sciences and Management Studies*, 10(4), 132–141. <https://doi.org/10.20448/ajssms.v10i4.5099>
- Zhai, X., & Razali, A. B. (2023). Triple method approach to development of a genre-based approach to teaching ESL/EFL writing: A systematic literature review by bibliometric, content, and scientometric analyses. *SAGE Open*, 13(1), 1–27. <https://doi.org/10.1177/21582440221147255>
- Zou, M., Kong, D., & Lee, I. (2021). Teacher engagement with online formative assessment in EFL writing during COVID-19 pandemic: The case of China. *Asia-Pacific Education Researcher*, 30(6), 487–498. <https://doi.org/10.1007/s40299-021-00593-7>