

Developing a Writing Assessment Instrument for English Students on Argumentative Essays

Suhono ^{1,2*}, Nur Mukminatien ¹, Erikson Saragih³, Raga Driyan Pratama ¹, Impiani Zagoto ⁴, Ardhi Eka Fadilah ¹, Risma Fahrul Amin ¹

¹ Universitas Negeri Malang, Indonesia

² Universitas Ma'arif Lampung, Indonesia

³ Universitas Sumatera Utara, Indonesia

⁴ Universitas Nias Raya, Indonesia

 suhono120708@gmail.com*

Abstract	Article Info
Writing argumentative essay poses a significant challenge for English students, who often struggle with expressing opinions, organizing ideas logically, and using appropriate language structures. Despite the importance of assessing writing skills at early stages, few reliable instruments exist that are specifically designed to evaluate argumentative essay. This study aims to develop a valid and reliable analytic rubric to assess students' argumentative essay skills, focusing on six key components: clarity of thesis, organization and coherence, supporting evidence, argument development, language use, and vocabulary range. The research employed a design-based development approach, incorporating literature review, expert validation, and empirical testing. Content validity was established through expert judgment (S-CVI = 0.95), and the instrument was pilot-tested with 85 English students. Statistical analyses demonstrated high inter-rater reliability (ICC = 0.87) and strong internal consistency (Cronbach's α = 0.89). Descriptive data revealed that while English students showed emerging abilities in thesis clarity and grammar, they struggled with supporting arguments and maintaining logical flow. The resulting rubric serves both as an assessment and instructional tool, helping educators diagnose English student needs and promote writing development. Unlike general-purpose rubrics or those designed for advanced English students, this rubric is specifically aligned with the challenges of early-stage academic writers. The study contributes to closing a gap in writing assessment tools and provides a practical resource for use in EFL classrooms.	Article History <i>Received :</i> <i>April 28, 2025</i> <i>Revised :</i> <i>May 28, 2025</i> <i>Accepted :</i> <i>June 05, 2025</i> Keywords: <i>Argumentative Writing, Writing Assessment, Analytic Rubric, Writing Skill Development,</i>
Published by ISSN Website This is an open access article under the CC BY SA license	Yayasan Payungi Smart Madani 2776-4303 https://journal.payungi.org/index.php/ijcep https://creativecommons.org/licenses/by-sa/4.0/



INTRODUCTION

Writing is a central component of language proficiency and academic success, particularly in contexts where English students are expected to express complex ideas, engage critically with content, and construct coherent arguments. Among the various genres of academic writing, argumentative essays are especially important as they require English students to articulate a position, support it with evidence, and address counterarguments in a structured and persuasive manner. However, for English students, mastering this genre poses significant challenges. Many struggles with articulating a clear thesis, organizing their ideas coherently, maintaining logical flow,

integrating supporting evidence, and adhering to the conventions of argumentative discourse. These challenges often stem not only from limited linguistic resources but also from underdeveloped critical thinking skills and unfamiliarity with formal academic writing structures. These difficulties often stem from limited vocabulary, underdeveloped critical thinking skills, and a lack of familiarity with the structure and conventions of argumentative texts. Despite these challenges, assessment tools specifically tailored to evaluate the argumentative essay abilities of English students in the first semester remain limited.

Despite the recognized importance of argumentative essay, there exists a notable lack of valid and reliable assessment tools specifically tailored to evaluate the argumentative essay abilities of novice English students. Most existing instruments are either designed for more advanced writers or focus on general writing ability rather than the distinctive features of argumentative texts. For instance, instruments such as the TOEFL iBT writing rubric or IELTS writing band descriptors (Fleckenstein et al., 2020; Liang, n.d.; Llosa & Malone, 2017), evaluate general academic writing skills but do not isolate or adequately weigh key components of argumentation, such as counterargument incorporation or logical structuring of evidence. Similarly, other widely used assessment rubrics in second language writing research tend to be too broad or are designed primarily for intermediate or advanced students ("Assess. Writ.," 2011; Hyland, 2019), making them less suitable for diagnosing specific challenges faced by English students.

Previous research has explored argumentative essay from a variety of angles, such as strategies employed by students (Nussbaum & Schraw, 2007), the effect of instructional interventions (Mayer, 2023), and the linguistic features of effective argumentation (Stapleton & Wu, 2015). However, these studies often assume a certain baseline proficiency and do not focus on assessment tool development for students in the first semester. For example, studies by (Ferris, 2004) and (Jo, 2021) examined error correction and discourse features in L2 writing but did not propose or validate assessment instruments tailored to entry-level students. Moreover, the available rubrics tend to emphasize surface-level language accuracy rather than deeper cognitive elements like reasoning, evidence, and logical flow components that are essential to argumentative essay.

In contrast, the aim of this study is to develop a valid and reliable assessment instrument specifically designed to evaluate the argumentative essay skills of English students, especially in English students in the first semester at Universitas Ma'arif Lampung and Institut Agama Islam Negeri Metro. The instrument will focus on key components of argumentative essays, such as clarity of thesis, coherence of argument, use of supporting evidence, and logical organization elements that are frequently overlooked in general-purpose rubrics. By targeting these foundational components, this instrument will fill an important gap in the literature and pedagogical practice.

The development process would be grounded in established writing theories and informed by models such as the Toulmin model of argumentation (Toulmin, 2003), which provides a framework for evaluating argument structure, and the Common European Framework of Reference for Languages (CEFR), which offers performance descriptors for various levels of language proficiency. The design will involve expert validation, pilot testing with target users, and rigorous statistical analysis to ensure the instrument's reliability and construct validity. This study therefore offers a novel contribution by focusing on the intersection of argumentative essay. While prior

instruments have addressed writing evaluation more generally, few have attempted to design developmentally appropriate, research-based tools that directly address the cognitive and rhetorical demands of argumentative essay for English students. By addressing this gap, the present study seeks to develop a valid and reliable analytic rubric to assess English students' argumentative essay skills, focusing on six key components: clarity of thesis, organization and coherence, supporting evidence, argument development, language use, and vocabulary range.

THEORETICAL REVIEW

Writing Skill Development in Language Learning

Writing is widely acknowledged as one of the most complex and demanding skills in language acquisition. It integrates various cognitive processes, including planning, organizing, drafting, and revising (Gregg & Steinberg, 2016; *Literacy: Major Themes in Education - Google Books*, n.d.). For English student in the first semester, writing presents considerable challenges, not only due to linguistic limitations but also due to underdeveloped rhetorical and metacognitive skills. Writing, especially in an academic context, demands the ability to structure ideas logically, use cohesive devices effectively, and tailor content to purpose and audience (Hyland, 2019). The development of writing skills in second language learners often follows a progression from simple sentence construction to more complex discursive forms, such as narrative and argumentative texts (Nation, 2009). This developmental trajectory necessitates instructional support and appropriate assessment tools that reflect English students' current proficiency levels and promote growth toward more advanced writing tasks.

Argumentative Essay as a Genre

Argumentative essay is considered a high-order genre within academic writing due to its requirement for logical reasoning, evidence-based support, and consideration of multiple perspectives. According to (Toulmin, 2003), model of argumentation, effective argumentative essay includes six key elements: claim, data, warrant, backing, qualifier, and rebuttal. This model serves as a foundational framework for evaluating the quality of arguments and is widely used in writing instruction and assessment. At the first semester, the English students often struggle with essential argumentative components such as clearly stating a thesis, organizing ideas logically, and using evidence persuasively (Nussbaum & Schraw, 2007). Studies have shown that explicit instruction in argument structures can improve English students' ability to formulate coherent arguments (Knudson, 1992). Therefore, assessing these components systematically is crucial for supporting early-stage development in argumentative essay.

Principles of Writing Assessment

Effective assessment of writing involves multiple dimensions, including content, organization, language use, mechanics, and task fulfillment. A valid assessment instrument should reflect both the theoretical constructs it aims to measure and the practical realities of English student performance (Mao & Lee, 2020). In designing writing rubrics, reliability (consistency of scoring) and validity (accuracy in measuring intended constructs) are essential considerations (Bachman & Adrian S. Palme, 1996). Rubrics and scoring instruments must align with the learning objectives and be developmentally appropriate for the target group. This means simplifying descriptors while still focusing on critical elements such as idea development, logical flow, and basic argumentative structure. Instruments designed for advanced or general populations

such as the IELTS Writing Band Descriptors or TOEFL iBT Rubrics may not sufficiently capture the nuanced needs of novice writers (Chalhoub-Deville, 2001).

The Role of Rubrics and Diagnostic Tools

Rubrics play a dual role in writing instruction: they serve as assessment tools and as instructional guides (Khairallah & Adra, 2022; Shabani & Panahi, 2020). Well-designed rubrics make expectations transparent, facilitate self-assessment and peer feedback, and enable teachers to deliver focused feedback. For English students' analytic rubrics, those that break down performance into discrete, assessable components are particularly useful (Moskal & Leydens, 2000).

Moreover, diagnostic assessment is especially critical at early stages of writing development. It provides insight into English students' strengths and areas needing improvement, which in turn informs curriculum design and instructional interventions (Alderson, 2005; Lee, 2015). In the context of argumentative essay, diagnostic tools should address core skills such as thesis development, paragraph organization, use of examples, and recognition of counterarguments.

Gaps in Existing Assessment Instruments

While numerous studies have focused on writing instruction and assessment, few have concentrated on developing specialized tools for assessing argumentative essay for English students. Most existing tools either assume intermediate to advanced proficiency or focus on holistic scoring without offering detailed diagnostic insights (Ferris, 2004; Kao et al., 2022; Ny & Doung, 2023). Additionally, many rubrics lack grounding in argumentation theory and fail to assess key dimensions such as counterargument inclusion, logical progression, or evidence integration. This gap underscores the need for an evidence-based, English students at the first semester appropriate instrument that can evaluate argumentative essay across cognitive, linguistic, and rhetorical domains.

METHOD

This study employed a Research and Development (R&D) design adapted from the model proposed by (Gall et al., 1996), which is widely used in educational research for developing instructional tools and assessment instruments. The research process followed a sequence of stages: needs analysis, theoretical foundation and rubric construction, expert validation, pilot testing, data analysis, and final product revision.

The target sample of the instrument are English students at the first semester, typically English students at Universitas Ma'arif Lampung and Institut Agama Islam Negeri Metro. The participants in the study included experts in academic writing and language assessment who provided feedback during the validation process, and 86 English students from university programs, selected using purposive sampling based on language proficiency. Additionally, 2 to 3 trained raters were involved in evaluating the English student writing samples using the rubric.

The data collection process consisted of several techniques. First, a needs analysis was conducted through document reviews and informal interviews with teachers to determine common challenges in assessing English students' argumentative essay. Based on findings and relevant literature, the initial rubric was developed. The rubric consisted of six key components: (1) clarity of thesis statement, (2) organization and coherence, (3) use of supporting evidence, (4) argument development, (5) language use and grammar, and (6) vocabulary range. Each component was rated on a 5-point Likert scale, with clearly defined descriptors for each level of performance. Expert

validation was conducted using structured evaluation sheets, and the Content Validity Index (CVI) was used to measure the relevance and clarity of each rubric item. Items with CVI scores below 0.78 were revised based on expert suggestions. Following validation, the rubric was pilot-tested by administering an argumentative essay prompt to the English student participants. The English students were asked to write a short essay (200–250 words) in response to a guided prompt under a 45-minute time limit. Their writing was independently scored by the trained raters using the rubric.

To ensure the reliability of the instrument, both internal consistency and inter-rater reliability were analyzed. Cronbach's Alpha was calculated to assess the internal consistency of the rubric, while Cohen's Kappa or Intraclass Correlation Coefficient (ICC) was used to evaluate agreement between raters. Descriptive statistics, such as mean scores and standard deviations, were also computed to analyze the performance trends across the rubric's components. All ethical considerations were carefully addressed throughout the research process. Informed consent was obtained from all participants, including English students, raters, and expert validators. English student work was anonymized to maintain confidentiality, and ethical clearance was secured from the relevant institutional review board. Through this rigorous methodological framework, the study aimed to produce a valid, reliable, and pedagogically practical instrument that could be used by teachers to more effectively assess and support the development of argumentative essay skills among students.

RESULT AND DISCUSSION

RESULT

In result section, the researcher presents the outcomes of the development, validation, and pilot testing of the writing skill assessment instrument designed for English students focusing on argumentative essays. The findings are reported in terms of content validity, inter-rater reliability, internal consistency, English student performance trends, and overall usability of the instrument.

Instrument Development Outcome

The final instrument was developed as an analytic rubric comprising six key components identified through literature synthesis and needs analysis: (1) clarity of thesis statement (Evaluates whether the main idea or argument is clearly stated and appropriately positioned in the text), (2) organization and coherence (Assesses the logical flow of ideas, including paragraph structure, transitions, and overall cohesion), (3) supporting evidence (Measures the use and integration of relevant examples, data, or citations to substantiate arguments), (4) argument development (Examines the depth and complexity of the argument, including counter-arguments and critical analysis), (5) language use and grammar (Looks at grammatical accuracy, sentence structure, and appropriate usage), and (6) vocabulary range (Evaluates the breadth and appropriateness of word choice in conveying meaning effectively).

Each component was assessed using a 5-point Likert scale with detailed descriptors for levels 1 (Beginning) through 5 (Proficient). The rubric was constructed to align with CEFR A1–A2 proficiency descriptors and theories of academic writing (Nation, 2008; Paltridge, 2014). This scale is useful for capturing various levels of student performance and allows for more nuanced feedback than binary or holistic assessments. Importantly, the rubric is aligned with the CEFR A1–A2 proficiency descriptors, which are internationally recognized benchmarks for language ability. This

alignment ensures that the assessment is not only suitable for internal evaluation but also comparable across different educational settings globally.

Expert Validation

The researcher involved four expert validators: Dr. Umi Yawisah, Dr. Dian Anggraini, Dr. Aria Septi Anggaira, and Dr. Agus Setiawan. These experts specialize in English education and assessment. They played a crucial role in the validation process of the rubric. Each validator assessed the relevance and clarity of the rubric components using a 4-point relevance scale, evaluating how well each element aligned with the criteria of being relevant, clear, and appropriate for English students. The validators systematically reviewed each rubric component to generate quantitative data, which was then analyzed using the Item-Level Content Validity Index (I-CVI) and the Scale-Level Content Validity Index (S-CVI). Four experts in English education and assessment validated the rubric for relevance, clarity. The validation involved four experts in English education and assessment, who reviewed the rubric components for relevance, clarity, and alignment. To evaluate each rubric component, the experts employed a 4-point relevance scale, and their evaluations were used to calculate the Item-level Content Validity Index (I-CVI) and the Scale-level Content Validity Index (S-CVI). Using a 4-point relevance scale, the Item-level Content Validity Index (I-CVI) for each component ranged from 0.88 to 1.00, with a Scale-level CVI (S-CVI) of 0.95, indicating excellent content validity (Lynn, 1986). The Item-level CVI (I-CVI) scores ranged from 0.88 to 1.00, demonstrating high agreement among the experts on the validity of the rubric components. Specifically, the components "Clarity of Thesis," "Organization & Coherence," and "Language Use & Grammar" received perfect scores of 1.00, suggesting unanimous agreement on their relevance and clarity. The component "Vocabulary Range" received a slightly lower but still high score of 0.94, indicating strong content validity. However, two components; "Supporting Evidence" and "Argument Development" received the lowest scores of 0.88. Minor revisions were made to descriptors for "Supporting Evidence" and "Argument Development" based on expert feedback. While still within the acceptable range (above 0.78 is generally acceptable for four raters), these lower scores signaled areas for potential improvement. As a result of the expert feedback, minor revisions were made to the wording of the descriptors for "Supporting Evidence" and "Argument Development." These revisions were aimed at improving alignment with the cognitive and language of English students, suggesting that the experts considered developmental appropriateness in their validation process. The revised descriptors were subsequently integrated into the final version of the rubric.

Table 1
Item-Level Content Validity Index (I-CVI) for Rubric Components

Component	I-CVI
Clarity of Thesis	1.00
Organization & Coherence	1.00
Supporting Evidence	0.88
Argument Development	0.88
Language Use & Grammar	1.00
Vocabulary Range	0.94

The "Item-Level Content Validity Index (I-CVI)" refers to a numerical measure that shows how much agreement there is among expert validators about the relevance of individual items (components) in a rubric. The I-CVI is a statistic used to evaluate how relevant and clear each item (or component) of an instrument (in this case, a writing rubric) is, based on expert judgment. Each expert rates every item on a 4-point scale of relevance: 1 = Not relevant, 2 = Somewhat relevant, 3 = Quite relevant and 4 = Highly relevant. The Item-Level Content Validity Index (I-CVI) is a statistical measure used to assess how much expert validators agree on the relevance and clarity of each individual component in a rubric. It is calculated by dividing the number of experts who rate an item as relevant (score 3 or 4 on a 4-point scale) by the total number of experts. An I-CVI score close to 1.00 indicates strong agreement and excellent content validity for that specific item.

The average S-CVI of the rubric was 0.95, indicating excellent content validity. Experts suggested minor wording improvements to two components ("Supporting Evidence" and "Argument Development") to better suit the cognitive level of English students. These suggestions were incorporated into the revised version. The Scale-level CVI (S-CVI) was calculated as 0.95, which is considered excellent by established benchmarks referencing (Lynn, 1986). This high score affirms that the rubric as a whole possesses a strong level of content validity. Written feedback was provided for components deemed less robust. For instance, the "Supporting Evidence" and "Argument Development" components received a lower I-CVI score of 0.88, prompting suggestions to revise their descriptions to better suit the abilities of English students. Based on the validators' input, the researcher revised the rubric accordingly. The research team made minor adjustments to several descriptors to ensure they aligned more closely with the pedagogical objectives.

In conclusion, the finding indicated that the rubric underwent a thorough validation process and emerged with strong content validity. The constructive feedback from domain experts led to meaningful refinements, especially in areas that required more clarity or better cognitive alignment. This process underscores the importance of expert involvement in educational assessment design and provides a reliable foundation for the rubric's application in real-world learning contexts.

Pilot Testing and Rater Reliability

The revised rubric was pilot-tested with 85 English students who completed an argumentative essays task based on a structured prompt. The students responded to an argumentative essays prompt under controlled classroom condition. The essays were evaluated by three trained raters. The pilot-tested with 85 English students refers to the practical implementation of the revised writing rubric in a real classroom setting, where 85 English students completed an argumentative essays task. The evaluation of their work by three trained raters, combined with statistical analyses such as the Intraclass Correlation Coefficient (ICC = 0.87) and Cronbach's Alpha ($\alpha = 0.89$), demonstrated both high inter-rater reliability and strong internal consistency. The involvement of trained raters is a critical aspect of rubric validation, as it minimizes variance due to subjective interpretation and increases the reliability of the assessment process. The training also ensures a common understanding of the rubric's components and scoring criteria, which is essential in evaluating inter-rater reliability.

To determine the degree of agreement among the raters, the study used the Intraclass Correlation Coefficient (ICC) for absolute agreement. The inter-rater reliability was calculated using Intraclass Correlation Coefficient (ICC) for absolute agreement. The Intraclass Correlation Coefficient (ICC) for the raters' scores was 0.87, reflecting excellent inter-rater reliability, based on standards proposed by (Koo & Li, 2016; Landers, 2023). To evaluate how well the six rubric components measured the same underlying construct, Cronbach's Alpha was calculated. Additionally, Cronbach's Alpha for the six rubric components was calculated to assess internal consistency, yielding a value of $\alpha = 0.89$, which indicates a high level of reliability. In other words, the rubric items are well-aligned with each other and reliably measure the same general domain of writing skills. This supports the notion that the rubric has strong psychometric properties and can be considered a valid tool for classroom-based assessment. This result confirms the rubric's inter-rater reliability and usability in classroom settings. This means that the raters consistently interpreted and applied the rubric in similar ways when scoring the English students' essays. A high ICC is crucial for confirming that a rubric is not only theoretically sound but also practically usable by different evaluators.

The high ICC and Cronbach's Alpha values provide empirical evidence of both inter-rater reliability and internal consistency, two critical dimensions of rubric quality. These findings suggest that the revised rubric is not only consistent in how different raters score the same piece of writing but also coherent in its internal structure. Consequently, the rubric is both usable and dependable for instructional use, providing teachers with a credible tool for evaluating and improving English students' argumentative essay skills.

The Descriptive Statistics of Student Performance

The descriptive statistics of student performance also provided insight into the instrument's practical use. The average total score was 17.4 (SD = 4.2) out of a maximum of 30 points. Descriptive analysis of student performance revealed moderate proficiency in "Thesis Clarity" and "Language Use," but lower scores in "Supporting Evidence" and "Argument Development." These results suggest that the students can articulate opinions but struggle to construct logical and well-supported arguments.

Table 2
Descriptive Statistics of Student Scores per Rubric Component (N = 85)

Component	Mean	SD
Thesis Clarity	3.4	0.8
Organization & Coherence	2.9	0.9
Supporting Evidence	2.6	1.0
Argument Development	2.5	0.9
Language Use & Grammar	3.2	0.8
Vocabulary Range	2.8	0.7

These scores reveal that students generally perform best in Thesis Clarity and Language Use & Grammar, suggesting that they are able to express opinions and construct grammatically correct sentences. However, the lower scores in Supporting Evidence and Argument Development indicate significant challenges in structuring logical arguments and backing them with relevant support key components of effective argumentative essay. These findings describe the specific areas that require pedagogical focus when teaching argumentative essay to students.

The rater feedback in post-assessment indicated that the rubric was easy to use, clear, and appropriate for the proficiency level of the students. Raters appreciated the detailed descriptors and suggested it be integrated into classroom writing instruction as both an assessment and instructional tool. Students also responded positively to the rubric's transparency, as it helped them understand the expectations of argumentative essay. The result reported that the rubric was clear, practical, and appropriate for students in the first semester, particularly due to the level-specific descriptors. Student feedback revealed improved awareness of writing expectations and increased confidence in constructing argumentative texts. The descriptors were deemed easy to use and aligned with student proficiency, making them suitable for integration into classroom instruction. Raters and students alike appreciated the rubric's transparency, which helped clarify expectations and boosted student confidence in understanding and constructing argumentative texts. These findings support the rubric's dual function as both an assessment and instructional tool. The developed rubric is a valid and reliable instrument for assessing argumentative essay. It demonstrated high content validity ($S-CVI = 0.95$), strong internal consistency ($\alpha = 0.89$), and excellent inter-rater reliability ($ICC = 0.87$).

These statistical outcomes confirm that the rubric is well-constructed and reliable for evaluating students' argumentative essay. The pilot study results also validated its diagnostic utility by identifying common writing challenges, such as underdeveloped arguments and weak use of evidence, thereby enabling more targeted instructional planning. Moreover, the results of the pilot study identified common writing challenges faced by novice writers and confirmed the rubric's diagnostic utility.

DISCUSSION

The central finding of this study is the successful development and validation of an analytic rubric that is both valid and reliable for assessing argumentative essay skills for students (CEFR A1–A2). The rubric demonstrated high content validity ($S-CVI = 0.95$), strong internal consistency (Cronbach's $\alpha = 0.89$), and excellent inter-rater reliability ($ICC = 0.87$). This confirms that the instrument is psychometrically sound and pedagogically appropriate for use in classroom settings.

While previous studies have explored rubrics for academic writing, most of them targeted intermediate to advanced students (e.g., B1–C2) or focused on general writing proficiency rather than specific genres like argumentative essay. For instance, ("Assess. Writ.," 2011; Smart, 2019) developed scoring instruments primarily used in large-scale assessments such as the TOEFL, which emphasize global writing competence and are often not tailored for classroom-based formative assessment. Similarly, (Bacha, 2010) examined argumentative essay development among university students but did not offer a level-specific instrument suitable for early-stage students.

In contrast, the present study makes a novel contribution the first by focusing on students in the first semester English Study Program, whose needs in constructing

argumentative texts are distinctly different from more proficient writers. At this stage, students struggle with expressing opinion, organizing ideas, and using logical connectors, all of which are addressed in the rubric's detailed descriptors. The second, developing a genre-specific rubric (for argumentative essay), as opposed to general writing rubrics found in many commercial frameworks (Jacobs & Others, 1981), which do not sufficiently emphasize argument structure, supporting evidence, or thesis clarity. The third, emphasizing developmental feedback, unlike summative-only rubrics. The rubric designed in this study includes level descriptors that help students understand progression and areas for improvement thereby serving a diagnostic and instructional function, in line with formative assessment best practices (Black & Wiliam, 2009). These distinctions address a critical gap identified by (Andrade & Du, 2005), who argued that many rubrics fail to support learning because they are too generic or too advanced for the students they assess.

The performance data from pilot testing revealed that English students performed best in areas such as thesis statement clarity and basic grammar usage, while consistently struggling with argument development and supporting evidence. This finding is consistent with the work of (Wingate, 2012), who noted that novice writers often have limited experience with the rhetorical conventions of argumentation. Therefore, teaching practices must place greater emphasis on helping students understand the structure of arguments, how to support claims with reasons or examples, and how to use appropriate discourse markers. Moreover, the rubric's clarity and ease of use were highlighted in feedback from raters, indicating that it can serve as a practical tool for both assessment and instruction. This dual functionality echoes the recommendation by (Panadero & Jonsson, 2013), who advocated for rubrics that support students' metacognitive development and writing self-efficacy. Unlike holistic scoring systems used in standardized tests such as the IELTS or TOEFL, which provide an overall band score, the current rubric is analytic and provides detailed, component-based feedback. While holistic scoring may be efficient for large-scale testing (CUMMING, 2011), analytic rubrics are superior for classroom teaching because they give students actionable insights into specific strengths and weaknesses. It is a key benefit especially for low-proficiency students.

The practical significance provides a targeted, level-appropriate tool that educators can use to assess and guide students in developing argumentative essay skills. The instrument supports lecturers by offering structured feedback and identifying specific areas where students need improvement. The developed instrument can inform syllabus and lesson planning, ensuring that instructional materials align with the specific needs and challenges of English students. This contributes to more effective teaching practices and improved student outcomes. The instrument may also serve as a self-assessment tool for students. English students can use it to independently monitor their writing progress, identify weaknesses, and revise their drafts more effectively. Then, educational institutions can adopt the instrument as a standardized tool to assess argumentative essay performance, helping to ensure consistency in evaluation and benchmarking across classrooms, schools, or even national programs.

The theoretical significance of this research deepens our understanding of how English students construct argumentative texts. It helps refine existing models of second language writing development by highlighting which aspects (e.g., coherence, reasoning, counter arguments) are most challenging for English students at the first

semester. The study contributes to the theoretical discourse around writing assessment by proposing a new or adapted framework that suits novice writers. It fills a gap in current assessment models that often do not adequately differentiate between levels of writing proficiency. The instrument developed can serve as a foundational tool for future empirical studies examining the progression of argumentative essay skills, instructional interventions, or comparative studies across proficiency levels and educational contexts.

Although the current study has produced a valid and reliable instrument, future research should expand the testing of the rubric across various educational contexts, including public and private schools, to examine its generalizability and effectiveness across different student populations. Long-term studies could investigate how the use of the rubric impacts students' development in argumentative essay over time. Such research would offer insights into the rubric's instructional value and its influence on learning outcomes. Given the increasing role of educational technology, researchers may explore the possibility of adapting the rubric into a digital format, potentially linked with automated writing evaluation systems. This could enhance efficiency and support remote learning environments. By addressing these areas, future research can continue to refine the tools used to support the writers and enhance the quality of language education globally. The present study serves as a foundation for such continued efforts, emphasizing the importance of level-appropriate, genre-specific assessment in fostering effective academic communication from the earliest stages of writing development.

CONCLUSION

Through a systematic development process including literature review, expert validation, pilot testing, and statistical analysis, a genre-specific analytic rubric was created. The rubric demonstrated high levels of content validity ($S-CVI = 0.95$), strong inter-rater reliability ($ICC = 0.87$), and excellent internal consistency (Cronbach's $\alpha = 0.89$), affirming its appropriateness for use in English students. The rubric focuses on six core components critical to argumentative essay: clarity of thesis, coherence and organization, use of supporting evidence, argument development, language use, and vocabulary range. Importantly, the rubric addresses a significant gap in existing research and practice, where most available writing assessment tools are either too general or tailored to intermediate-to-advanced students. This study thus contributes not only to the field of language assessment but also to classroom-based pedagogy by providing a practical tool that promotes both evaluation and instructional support.

Based on the findings of this study, several practical suggestions can be made for educators and practitioners. The rubric should be integrated into classroom teaching to enhance transparency in writing expectations and provide clear, structured feedback to students. This can help students understand the components of effective argumentation and develop metacognitive awareness about their writing processes. For optimal use, teachers and evaluators should be trained to use the rubric consistently. A rubric is only as effective as its application, and inter-rater reliability can be improved through collaborative norming sessions. Teachers are encouraged to introduce the rubric to students as a self-assessment and peer-assessment tool. Doing so can foster student autonomy and reflective learning, which are essential for writing development, especially at the first semester of English students.

REFERENCES

- Alderson, J. C. (2005). Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment. *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*, 1–284. <https://doi.org/10.1080/15434300701595637/ASSET//CMS/ASSET/1B727B33-AC2B-4A2A-94D8-03055FE3A940/15434300701595637.FP.PNG>
- Andrade, H., & Du, Y. (2005). Student Perspectives on Rubric-Referenced Assessment. *Educational & Counseling Psychology Faculty Scholarship*. https://scholarsarchive.library.albany.edu/edpsych_fac_scholar/2
- Assessing Writing. (2011). *Concepts in Composition*, 159–194. <https://doi.org/10.4324/9780203806807-9>
- Bacha, N. N. (2010). Teaching the academic argument in a university EFL environment. *Journal of English for Academic Purposes*, 9(3), 229–241. <https://doi.org/10.1016/J.JEAP.2010.05.001>
- Bachman, L., & Adrian S. Palme. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press. https://books.google.co.id/books?hl=en&lr=&id=E0yH0NdySrQC&oi=fnd&pg=PP9&dq=anguage+Testing+in+Practice:+Designing+and+Developing+Useful+Language+Tests&ots=CEzZeEKLGG&sig=waCakouIqXYqAvGq8w5UEbnvLVg&redir_esc=y#v=onepage&q=anguage+Testing+in+Practice%3A+Designing+and+Developing+Useful+Language+Tests&f=false
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/S11092-008-9068-5/FIGURES/2>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press. https://books.google.co.id/books?hl=en&lr=&id=PygQ8Gk4k4YC&oi=fnd&pg=PR9&dq=ommon+European+Framework+of+Reference+for+Languages:+Learning,+Teaching,+Assessment&ots=SjrQf2zn7h&sig=rPQywcXNJe2ocPtElGL0qaWR0-M&redir_esc=y#v=onepage&q=ommon+European+Framework+of+Reference+for+Languages%3A+Learning%2C+Teaching%2C+Assessment&f=false
- CUMMING, A. (2011). Diagnostic Writing Assessment: The Development and Validation of a Rating Scale by KNOCH, UTE. *The Modern Language Journal*, 95(3), 476–477. https://doi.org/10.1111/J.1540-4781.2011.01212_19.X
- Ferris, D. R. (2004). The “Grammar Correction” Debate in L2 Writing: Where are we, and where do we go from here? (and what do we do in the meantime ...?). *Journal of Second Language Writing*, 13(1), 49–62. <https://doi.org/10.1016/J.JSLW.2004.04.005>
- Fleckenstein, J., Keller, S., Krüger, M., Tannenbaum, R. J., & Köller, O. (2020). Linking TOEFL iBT® writing rubrics to CEFR levels: Cut scores and validity evidence from a standard setting study. *Assessing Writing*, 43, 100420. <https://doi.org/10.1016/J.ASW.2019.100420>
- Gall, M. D., Walter R. Borg, & Joyce P. Gall. (1996). Educational research: An introduction, 6th ed. *Longman Publishing*. <https://psycnet.apa.org/record/1996-97171-000>
- Gregg, L. W. ., & Steinberg, E. R. (2016). *A Framework for a Cognitive Theory of Writing*. 51–72. <https://doi.org/10.4324/9781315630274-5>
- Hyland, K. (2019). *Second language writing*. Cambridge university press.

- https://books.google.co.id/books?hl=en&lr=&id=6DCdDwAAQBAJ&oi=fnd&pg=PR11&dq=Second+language+writing&ots=4nCwHjOx90&sig=q98-3Dg8FSUyCHJQEAPiunDrDw&redir_esc=y#v=onepage&q=Second language writing&f=false
- Jacobs, H. L., & Others, A. (1981). *Testing ESL Composition: A Practical Approach. English Composition Program*. 139.
- Jo, C. W. (2021). Short vs. extended adolescent academic writing: A cross-genre analysis of writing skills in written definitions and persuasive essays. *Journal of English for Academic Purposes*, 53, 101014. <https://doi.org/10.1016/J.JEAP.2021.101014>
- Kao, C. W., Reynolds, B. L., & Feng Teng, M. (2022). What we need to know about student writers' grammar learning and correction. *Applied Linguistics Review*, 13(2), 175–199. <https://doi.org/10.1515/APPLIREV-2019-0016/MACHINEREADABLECITATION/RIS>
- Khairallah, M., & Adra, O. (2022). The multifaceted function of rubrics as formative assessment tools: A classroom-based action research in an L2 writing context. *Language Teaching Research*. <https://doi.org/10.1177/13621688221104210>
- Knudson, R. E. (1992). Analysis of Argumentative Writing at Two Grade Levels. *The Journal of Educational Research*, 85(3), 169–179. <https://doi.org/10.1080/00220671.1992.9944434>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/J.JCM.2016.02.012>
- Landers, R. (2023). Computing Intraclass Correlations (ICC) as Estimates of Interrater Reliability in SPSS. *Authorea Preprints*. <https://doi.org/10.15200/WINN.143518.81744>
- Lee, Y. W. (2015). Diagnosing diagnostic language assessment. *Language Testing*, 32(3), 299–316. <https://doi.org/10.1177/0265532214565387>
- Liang, S. (n.d.). Documenting the Task Response Features of IELTS Writing Task 2, Band Scores 5-7 to Improve Teaching and Learning for Chinese Students Who Use IELTS Preparation Templates. *The Educational Review, USA*, 2024(3), 355–362. <https://doi.org/10.26855/er.2024.03.003>
- Literacy: Major Themes in Education - Google Books*. (n.d.). Retrieved May 23, 2025, from https://books.google.co.id/books?hl=en&lr=&id=eFdrzKWyl8IC&oi=fnd&pg=PA40&dq=+cognitive+process+theory+of+writing&ots=_OL9oxoSz1&sig=FeWNmMZl10BRFFe7GhUPXovubOM&redir_esc=y#v=onepage&q=cognitive process theory of writing&f=false
- Llosa, L., & Malone, M. E. (2017). Student and instructor perceptions of writing tasks and performance on TOEFL iBT versus university writing courses. *Assessing Writing*, 34, 88–99. <https://doi.org/10.1016/J.ASW.2017.09.004>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–386. https://journals.lww.com/nursingresearchonline/citation/1986/11000/determination_and_quantification_of_content.17.aspx
- Mao, Z., & Lee, I. (2020). Feedback scope in written corrective feedback: Analysis of empirical research in L2 contexts. *Assessing Writing*, 45, 100469. <https://doi.org/10.1016/J.ASW.2020.100469>
- Mayer, R. E. (2023). How to Assess Whether an Instructional Intervention Has an Effect on Learning. *Educational Psychology Review*, 35(2), 1–6.

- <https://doi.org/10.1007/S10648-023-09783-9/TABLES/1>
- Moskal, B. M., & Leydens, J. A. (2000). Scoring Rubric Development: Validity and Reliability. *Practical Assessment, Research, and Evaluation*, 7(1). <https://doi.org/10.7275/Q7RM-GG74>
- Nation, I. S. P. (2008). Teaching ESL/EFL Reading and Writing. *Teaching ESL/EFL Reading and Writing*. <https://doi.org/10.4324/9780203891643>
- Nussbaum, E. M., & Schraw, G. (2007). Promoting Argument-Counterargument Integration in Students' Writing. *The Journal of Experimental Education*, 76(1), 59–92. <https://doi.org/10.3200/JEXE.76.1.59-92>
- Ny, S., & Doung, D. (2023). Examining debates on grammar correction in second language writing classes: Perspectives, challenges, and implications. *Journal of Language, Literature, Social and Cultural Studies*, 1(2), 37–43. <https://doi.org/10.58881/JLLSCS.V1I2.38>
- Paltridge, B. (2014). Genre and second-language academic writing. *Language Teaching*, 47(3), 303–318. <https://doi.org/10.1017/S0261444814000068>
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/J.EDUREV.2013.01.002>
- Shabani, E. A., & Panahi, J. (2020). Examining consistency among different rubrics for assessing writing. *Language Testing in Asia*, 10(1), 1–25. <https://doi.org/10.1186/S40468-020-00111-4/TABLES/10>
- Smart, J. (2019). Affordances of TOEFL writing tasks beyond university admissions. *Assessing Writing*, 41, 80–83. <https://doi.org/10.1016/J.ASW.2019.06.006>
- Stapleton, P., & Wu, Y. (Amy). (2015). Assessing the quality of arguments in students' persuasive writing: A case study analyzing the relationship between surface structure and substance. *Journal of English for Academic Purposes*, 17, 12–23. <https://doi.org/10.1016/J.JEAP.2014.11.006>
- Toulmin, S. (2003). *The uses of argument*. Cambridge university press. https://books.google.co.id/books?hl=en&lr=&id=8UYgegaB1S0C&oi=fnd&pg=PR7&dq=The+Uses+of+Argument&ots=Xg24ljGSwS&sig=y4Y9bFQ5w466F2e-hKkvSkqhSD0&redir_esc=y#v=onepage&q=The Uses of Argument&f=false
- Wingate, U. (2012). 'Argument!' helping students understand what essay writing is about. *Journal of English for Academic Purposes*, 11(2), 145–154. <https://doi.org/10.1016/J.JEAP.2011.11.001>

Copyright holder:

© Suhono et al., (2025)

First Publication Right:

International Journal of Community Engagement Payungi

This article is licensed under:

CC-BY-SA

