

The Developing a Valid and Reliable Assessment Instrument for Academic Essay Writing in Higher Education

M. Hilmy Hidayatullah ^{1*}, Nadrotin Mawaddah ¹, Nur Mukminatien ², Suhono ³, Mukminatius Zuhriyah ², Nahid Ayad ⁴, Yeasy Agustina Sari ⁵


¹ Universitas Ibrahimy, Indonesia

² Universitas Negeri Malang, Indonesia

³ Universitas Ma'arif Lampung, Indonesia

⁴ University of Zawia, Libya

⁵ Universitas Islam Negeri Jurai Siwo Lampung, Indonesia

 ibnua7631@gmail.com*

Abstract

Many university instructors assess academic essay writing using general rubrics or borrowed instruments that do not align with specific course outcomes and learner profiles. This study was initiated to address this gap by developing a valid and reliable assessment instrument for academic essay writing among EFL university students, focusing on argumentative essays. The instrument was constructed systematically, involving needs analysis, expert validation, pilot testing, and statistical evaluation for reliability and validity. The trial was conducted with 6 students from the English Language Education Study Program, and their essays were rated by two trained raters. The results showed that the rubric successfully captured performance variation across six key writing components and allowed for consistent scoring with minimal discrepancies between raters. Expert validators confirmed that the rubric is clear, relevant, and aligned with academic essay writing expectations. Rater reliability analysis showed consistent inter-rater scores (score differences ≤ 1 point), indicating strong rubric usability. Based on these findings, this instrument offers a practical and theoretically grounded tool for assessing academic essay writing in EFL higher education settings.

Keywords: Academic Essay Writing Assessment, Instrument Development, Validity, Reliability, Higher Education

ARTICLE INFO

Article history:

Received
January 09, 2025

Revised
April 14, 2025

Accepted
April 30, 2025

Published by
ISSN

Website

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

CV. Creative Tugu Pena
2774-4299

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

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



INTRODUCTION

Academic essay writing plays a central role in higher education, especially in English as a Foreign Language (EFL), where it serves as a learning tool and a measure of academic achievement. Both are complex tasks because to obtain a proper academic paper, students must possess several skills, such as creative thinking, language delivery, and idea construction (Elander et al., 2006; Fatmawati et al., 2024; Madkur, 2024; Yuskar, 2024). Over time, academic literacy has evolved from one generation to the next. Teaching academic essay writing to each

generation requires upgrading skills related to digital and technological transformation (Andheska et al., 2020; Polakova & Ivenz, 2024; Wulandari et al., 2023). Furthermore, academic essay writing in English as a foreign language has more complicated challenges regarding translingual issues (Gustafsson & Ganobcsik-Williams, 2023). The complexity of the teaching process results in intricacy in assessing the outcomes as it consists of several criteria such as internal, coherence, and infirmity (Pipalova, 2024). Those challenges influenced further research to explore the field of assessment instruments on students' academic essay writing.

Effective writing assessment requires more than a checklist of errors. It must be valid, reliable, and practical. That means measuring the skills we want to teach, producing consistent results across raters, and being usable in the classroom without excessive burden (Barkaoui, 2010; Brown & Abeywickrama, 2010). Analytic rubrics are a promising solution to these challenges because they offer detailed, criterion-based scoring and can prominently guarantee validity and reliability (Yeo et al., 2024). However, even though these rubrics must be designed carefully, without alignment to course goals, explicit descriptors, and local validation, they risk becoming another one-size-fits-all tool that does not serve learners effectively. Those are the requirements that instructors must consider when developing assessment instruments for academic essay writing.

Despite the standard, it should be noted that not all lecturers adhere systematically to these requirements. Indeed, there is considerable misalignment among lecturers regarding criteria, rubrics, and instruments, raising concerns regarding the assessment process's reliability and validity (Clarkson, 2024; Mukminatien et al., 2025; Rizqi & Faujianor, 2024). These instruments tend to emphasize surface-level correctness, such as grammar and punctuation, while giving less attention to deeper aspects of writing, including idea development and academic integrity (East & Slomp, 2024). As a result, the feedback students receive may be unclear, unreliable, and misaligned with learning outcomes.

In the Indonesian context, the issue is compounded by the lack of writing assessment instruments explicitly developed for local university classrooms. Although many institutions offer academic essay writing courses, few provide rubrics grounded in the tasks students are asked to perform or the standards they are expected to meet. Based on preliminary classroom observations, many lecturers continue to use improvised rubrics or ones adapted from international tests, with limited reference to the specific features of argumentative writing. This disconnection poses challenges for fair assessment and effective instruction.

Based on the preliminary study, it was also found that there is an aspect that these lecturers often neglect to consider is the type or genre of text to be assessed. For instance, the rubrics and criteria used to assess argumentative essays were the same as the assessment of other types of academic essay writing. It is essential to consider that the evaluation of argumentative essays should prioritize students' ability to articulate their arguments and organize them with evidence in an essay (Iordanou & Rapanta, 2021). For this reason, critical thinking is paramount in the assessment process (Sato, 2022). Moreover, this genre is characterized by its unique generic structures: an introduction to argue a position, body paragraphs to explain reasons, a counterargument, and a conclusion (Yang,

2022). It involves cognitive, linguistic, and psychological aspects, such as fear of making mistakes, low self-esteem, and low motivation (Salihoglu, 2024). Therefore, assessing students' argumentative texts requires instruments designed meticulously for this text genre.

The previous paragraphs have explained the gap between the theoretical requirements in academic essay writing assessment and the factual practices at the higher education level. Therefore, this study aims to fill the gap by developing a valid and reliable classroom-based assessment instrument to assess university students' performance in writing argumentative essays in English. The instrument consists of a task prompt and an analytic scoring rubric that captures key aspects of academic essay writing, including task achievement, coherence, argumentation, language accuracy, academic style, and mechanics. The rubric was constructed based on a thorough literature review, expert consultation, and classroom testing to ensure that it reflects both theoretical principles and practical realities.

The significance of this study lies in its attempt to bridge theory and practice. While previous research has provided theoretical insights into writing assessment (Chauhan, 2022; Dong et al., 2024; East, 2009; Hyland, 2013; Knoch, 2011; Zheng & Yu, 2019), much of it focuses on large-scale standardized tests or general-purpose rubrics. The current trend is toward corrective feedback strategies in formative and summative writing assessments (Morris et al., 2021; Olsen & Hunnes, 2024) rather than going back to basics to maintain the basic criteria for obtaining authentic evaluation. There remains a lack of research on instruments tailored for classroom use in EFL contexts, especially within curriculum policies like Indonesia's Merdeka Belajar, which emphasizes critical thinking, contextual relevance, and competency-based learning (Kemdikbud, 2020).

What this study offers, therefore, is not just another rubric but a locally grounded, empirically tested instrument that is sensitive to the genre of argumentative writing and the pedagogical goals of higher education writing instruction. The novelty of this research lies in its dual focus: aligning the rubric with genre-based writing pedagogy while subjecting it to systematic validation involving expert review and empirical testing with actual student writing.

The study is guided by the hypothesis that a well-constructed analytic rubric—developed through expert validation and piloted in real classroom conditions—can yield reliable scoring outcomes and offer meaningful feedback to both students and instructors. Six writing components are used as variables for assessment: task achievement, organization and coherence, argument and evidence, language use and accuracy, academic style, and mechanics. The research follows a developmental design involving needs analysis, instrument construction, expert validation, rater calibration, and statistical analysis of inter-rater reliability.

Throughout the paper, a few terms will recur: academic essay writing refers specifically to argumentative essays written in formal academic English; rubric denotes a set of scoring criteria used to assess student work; and inter-rater reliability refers to the degree of agreement among scorers. These terms will be used consistently to describe the processes and findings related to the assessment instrument.

This study is to develop a valid, reliable, and contextually relevant performance-based assessment instrument to evaluate argumentative essay writing among university-level EFL students. The focus is on creating a rubric that is not only valid and reliable, but also practical for classroom use and aligned with the actual objectives of academic writing courses. Through expert review and try outs, the study seeks to ensure that the instrument reflects key aspects of academic writing such as clarity of argument, coherence, language use, and adherence to academic conventions, while also supporting teachers in providing more consistent and meaningful feedback.

METHOD

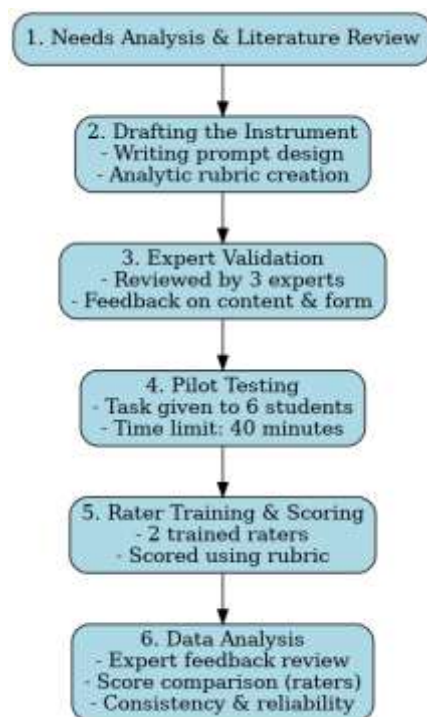
This study used a developmental research approach to design and test an assessment tool for evaluating argumentative essay writing among EFL university students. The goal was to create a rubric that is not only valid and reliable, but also practical for classroom use. To achieve this, the research went through several stages, including identifying classroom needs, designing the rubric, getting feedback from experts, trying it out with students, and checking the consistency of scores between raters. Each step was aimed at making sure the instrument matched the goals of academic writing courses and could be used effectively by teachers in real classroom settings.

Research Design

This study employed a developmental research design. It aimed to construct, validate, and evaluate a writing assessment instrument for academic argumentative essays in higher education EFL contexts. The model adopted follows the key phases of instrument development proposed by (Brown, 2004) and adapted by (Brachman & Palmer, 2010), including needs analysis, instrument design, expert validation, pilot testing, rater scoring, and psychometric evaluation (validity and reliability testing).

The development process began with a needs analysis to identify gaps in current assessment practices and to ensure alignment with both course outcomes. It was followed by a systematic construction phase, which included designing the writing task and developing an analytic rubric tailored to six core components of academic essay writing. Validation of the instrument involved three expert reviewers in writing assessment and ELT, while a pilot implementation phase engaged six EFL students whose essays were scored independently by two trained raters. Data from the pilot were analyzed to assess the instrument's practicality, inter-rater reliability, and alignment with academic essay writing standards. The whole procedure of instrument development is illustrated in Figure 1.

Figure 1. Stages of Developing the Academic Essay Writing Assessment Instrument



The development of the academic essay writing assessment instrument followed a structured, multi-phase process designed to ensure its validity, reliability, and practical usability in an EFL university classroom. The process began with a needs analysis and literature review. At this stage, the current classroom practices and existing writing assessment instruments were examined, revealing a clear gap in localized and criterion-based tools aligned with Indonesian higher education contexts. Relevant literature on language testing, rubric design, and EFL writing assessment was also reviewed to inform the theoretical framework and guide the rubric structure (Brown & Abeywickrama, 2010; Hyland, 2013; Weigle, 2002).

Based on these findings, the second step involved drafting the assessment instrument. It included designing a writing task requiring students to produce an argumentative essay of 500–700 words within 90 minutes. The task was contextualized to match typical university-level expectations and writing genres. Simultaneously, an analytic scoring rubric was constructed, comprising six assessment components: (1) Task Achievement, (2) Organization and Coherence, (3) Argument and Evidence, (4) Language Use and Accuracy, (5) Academic Style (the use of formal tone, objective language, and adherence to academic conventions in writing), and (6) Mechanics. Each component was scored on a five-point scale, with clearly defined descriptors ranging from “Poor” (1) to “Excellent” (5). The rubric aimed to balance linguistic accuracy with cognitive and rhetorical elements of academic essay writing.

The third phase was expert validation. The draft rubric and task prompt were submitted to three applied linguistics and language assessment experts. These validators reviewed the instrument using a structured validation form, assessing aspects such as instructional clarity, task relevance, alignment with academic essay writing standards, rubric coverage, and practical feasibility. Their feedback informed several minor revisions, such as clarifying rubric descriptors and emphasizing the role of counterarguments in the task guidelines.

Next, the instrument entered the pilot testing phase, where six EFL university students completed the writing task under controlled classroom conditions. These participants were enrolled in an Academic Essay Writing course and taught in argumentative essay writing, ensuring the task was appropriate for their skill level. The written outputs were then anonymized and submitted for scoring.

The fifth step was rater training and independent scoring. Two experienced writing assessors were trained to use the rubric consistently, including guided calibration sessions using sample texts. Each rater then independently scored the student essays using the rubric. To ensure scoring accuracy and fairness, raters were not informed of each other's scores during the process.

The final step involved data analysis. The expert validation forms were reviewed, identifying strengths and suggestions for rubric clarity, construct alignment, and scoring practicality. Essay scores from both raters were compared to determine inter-rater reliability, with attention to whether score differences exceeded one point per criterion—a threshold commonly used in writing assessment research (Jonsson & Svingby, 2007). These analyses' results helped establish the instrument's initial effectiveness and consistency as a classroom-based academic essay writing assessment tool.

Data Collection and Analysis

The data collection process in this study encompassed three primary sources: expert validation forms, student writing samples, and rater score sheets. First, to assess content validity, three expert validators—specialists in applied linguistics and language assessment—were asked to complete structured validation forms after reviewing the task prompt and the analytic rubric. These forms included both quantitative ratings and qualitative comments regarding the clarity of the instructions, the task's relevance, the rubric's alignment with academic essay writing conventions, and the practicality of implementation. The responses provided critical insights into the strengths and potential areas for improvement of the instrument.

Second, student writing samples from six undergraduate EFL students enrolled in an Academic essay Writing course were collected. Each student completed the argumentative essay task under timed conditions (90 minutes) and followed the given academic essay writing instructions. These essays served as the core data for assessing the instrument's practical application and evaluating scoring consistency.

Third, score sheets from two trained raters who independently assessed the student essays were gathered using the developed rubric. The raters provided detailed scores across six criteria for each essay, and their assessments were conducted independently to ensure objectivity. In addition to numerical scores,

raters were encouraged to write brief qualitative comments explaining their judgments, which added depth to the evaluation and supported the interpretation of score patterns.

The data analysis began with a descriptive review of the expert validation forms to summarize agreement levels and identify any recurring feedback across validators. Suggestions regarding rubric clarity, alignment, and item specificity were synthesized to justify minor rubric revisions. Following this, the rater scoring data were analyzed for inter-rater reliability by comparing score differences across all six criteria for each student. As suggested by Jonsson and Svingby (Jonsson & Svingby, 2007), an acceptable threshold of agreement was set at a maximum difference of one point per criterion between raters. The consistency of scoring was measured using basic descriptive statistics, focusing on the frequency and extent of agreement or discrepancy. The combination of validation data and scoring consistency provided a robust basis for determining the instrument’s reliability, usability, and potential for classroom integration.

RESULT AND DISCUSSION

Students’ Performance Based on Rubric Criteria

The developed assessment instrument was piloted with six undergraduate students enrolled in an Academic Essay Writing course. Each participant completed an argumentative essay: *“Should university students be required to take general education courses outside their major?”* Their essays were evaluated using an analytic rubric comprising six criteria: task achievement, organization and coherence, argument and evidence, language use and accuracy, academic style, and mechanics. Each criterion was scored on a scale of 1 (Poor), 2 (Weak), 3 (Fair), 4 (Good) and 5 (Excellent, resulting in a total possible score range of 6 to 30 per essay. Table 1 presents the descriptive statistics for student scores.

Table 1. Descriptive Statistics of Student Performance (N = 6)

Criterion	Description	Mean (M)	Standard Deviation (SD)	Min	Max
Task Achievement	Relevance to the prompt and development of clear, well-supported arguments	2.83	0.75	2	4
Organization & Coherence	Logical flow of ideas, effective paragraphing, and transitions between points	2.83	0.41	2	3
Argument & Evidence	Strength of claims, use of relevant and convincing supporting details or examples	3.00	0.63	2	4
Language Use & Accuracy	Control of grammar and vocabulary, clarity and complexity of language	2.50	0.55	2	3
Academic Style	Use of formal register, tone, and adherence to academic writing	2.33	0.52	2	3

Criterion	Description	Mean (M)	Standard Deviation (SD)	Min	Max
	conventions				
Mechanics	Correctness in spelling, punctuation, and formatting	2.50	0.55	2	3

The highest average score was observed in Argument and Evidence (M = 3.00), indicating that students were relatively successful in presenting opinions and supporting them with reasons. It suggests a moderate understanding of argumentative structure. However, lower scores in Academic Style (M = 2.33) and Language Use (M = 2.50) suggest challenges in using formal registers and controlling grammatical accuracy. It aligns with known issues in EFL academic essay writing, where students often struggle with syntactic complexity and stylistic conventions (Hyland, 2007; Weigle, 2002).

Rater Agreement and Inter-Rater Reliability

To examine the instrument's reliability, both raters independently scored the essays without consultation. Each rater assigned a score to all six criteria, resulting in 36 data points (6 essays × 6 criteria).

Table 2. Rater Score Differences and Consistency

Student	Rater 1 Total	Rater 2 Total	Score Difference	Consistency
A	12	12	0	Perfect
B	17	18	1	Acceptable
C	20	21	1	Acceptable
D	16	16	0	Perfect
E	15	14	1	Acceptable
F	16	15	1	Acceptable

The results show a high level of inter-rater consistency. No student had more than a one-point difference between raters, which meets the acceptable threshold recommended by Jonsson and Svingby (Jonsson & Svingby, 2007). These findings confirm the rubric's clarity and practical usability in producing consistent scores across evaluators.

Expert Validation Ratings

To assess the content and construct validity of the rubric, three academic essay writing experts were consulted. Each expert assessed the six rubric criteria using a 5-point Likert scale (1 = Not valid, 5 = Highly valid). Table 3 summarizes the average expert ratings per criterion.

Table 3. Expert Validator Scores on Rubric Validity (1–5 scale)

Criterion	Expert 1	Expert 2	Expert 3	Mean
Task Achievement	5	5	5	5.00

Criterion	Expert 1	Expert 2	Expert 3	Mean
Organization & Coherence	4	5	5	4.67
Argument & Evidence	5	5	5	5.00
Language Use & Accuracy	4	4	5	4.33
Academic Style	4	5	4	4.33
Mechanics	4	4	5	4.33

All components received high ratings, particularly Task Achievement and Argument & Evidence, confirming their alignment with academic essay writing standards. Minor suggestions for Academic Style and Mechanics were made, encouraging further clarification between adjacent score levels. Nevertheless, all criteria were deemed valid and pedagogically appropriate.

DISCUSSION

The primary aim of this study was to design, validate, and evaluate a performance-based assessment instrument for academic argumentative essay writing in an EFL higher education context. The findings from expert validation, pilot testing, and inter-rater analysis collectively confirm that the instrument is both psychometrically sound and pedagogically aligned with the goals of academic essay writing instruction.

The highest mean score across the six rubric criteria was observed in Argument and Evidence ($M = 3.00$), indicating that students were relatively proficient in presenting positions and justifying them logically. This result suggests that classroom instruction on the claim-support structure was effective, particularly in teaching students how to structure arguments—a core feature of academic literacy (Hyland, 2013; Iordanou & Rapanta, 2021; Nussbaum & Schraw, 2007). Additionally, moderate scores in Task Achievement and Organization and Coherence (both $M = 2.83$) reflect students' developing ability to address prompts clearly and structure essays logically. These findings imply that students are beginning to internalize genre conventions of academic argumentation.

By contrast, the lowest scores appeared in Academic Style ($M = 2.33$) and Language Use and Accuracy ($M = 2.50$), suggesting that students continue to struggle with formal register, grammatical complexity, and lexical precision—common issues in EFL writing (Ehsanzadeh & Dehnad, 2024; Ferris & Hedgcock, 2023; Weigle, 2002). These weaknesses are consistent with previous studies that have noted persistent difficulties among EFL learners in maintaining grammatical accuracy and stylistic appropriateness under timed writing conditions (Barkaoui, 2010; Wubalem, 2021). Similarly, challenges in Mechanics ($M = 2.50$) point to the need for continued instruction in punctuation and spelling—subskills often underestimated in academic essay writing instruction but critical for clarity and fluency.

Expert validation of the rubric further supported its content and construct alignment. All six criteria were rated between 4.33 and 5.00 on a 5-point scale by three experienced language assessors. In particular, *Task Achievement* and *Argument and Evidence* were unanimously rated highly valid ($M = 5.00$), confirming that these criteria accurately reflect core competencies in

argumentative writing. Slightly lower scores in *Academic Style* and *Mechanics* ($M = 4.33$) suggest that these areas may benefit from more refined descriptors to clearly distinguish performance levels. This is consistent with the view that ambiguous or overlapping descriptors can reduce scoring accuracy and formative precise evaluation (Alemdag & Narciss, 2025; Asli et al., 2024).

The inter-rater reliability analysis demonstrated excellent consistency between two trained assessors. No scoring difference exceeded one point, and the Cohen's Kappa score of 0.800 indicates strong agreement, affirming the rubric's clarity and the effectiveness of rater training. These results meet the standards of classroom-based writing assessment, where transparent and consistent evaluation is essential for both summative grading and formative feedback (Brookhart, 2013; Brown & Abeywickrama, 2010; Morris et al., 2021; Zheng & Yu, 2019).

Simultaneously, the findings support the instrument's use as a practical tool for assessing argumentative writing in EFL classrooms. It successfully identifies both student strengths (e.g., idea development and argumentation) and areas for improvement (e.g., formal style and mechanics), making it a valuable diagnostic and instructional aid. Moreover, the instrument's alignment with the *Merdeka Belajar* Curriculum, which emphasizes critical thinking and communication, enhances its relevance to Indonesia's current educational policy context.

However, the study is not without limitations. The small sample size ($N = 6$) and limited diversity of student proficiency levels restrict the generalizability of the findings. Additionally, while expert validation was comprehensive, broader triangulation involving student and instructor feedback could further enrich the validation process. Future studies should expand the pilot to include students across different institutions and writing proficiencies and conduct longitudinal research to assess the instrument's impact on writing improvement over time.

CONCLUSION

This study set out to develop an assessment tool that could genuinely support the way EFL students are evaluated in writing argumentative essays. The result is a rubric that not only reflects the key aspects of academic writing but also works well in real classroom settings, with clear criteria and consistent scoring. The entire process was shaped by practical teaching experiences and feedback from actual users, making the instrument both relevant and usable for teachers and students alike.

Pilot results indicated that students performed most strongly in developing arguments and supporting them with evidence, while challenges remained in achieving grammatical accuracy, maintaining academic tone, and controlling mechanics. These performance trends suggest that the instrument effectively distinguishes skill areas, providing instructors with actionable insights into student writing abilities. Expert validation confirmed strong content and construct alignment of the rubric with academic essay writing standards, and inter-rater reliability analysis demonstrated high-scoring consistency between assessors. Concurrently, these findings affirm the instrument's validity, reliability, and practicality for classroom use. Theoretically, this study contributes to the field of EFL writing assessment by offering a replicable model grounded in genre-based pedagogy and test development principles. Practically, the rubric serves as a tool

not only for grading but also for delivering formative feedback and guiding instructional planning. It would be beneficial for future research to involve larger and more diverse student groups, integrate learner feedback to assess task clarity, and investigate the potential of digital platforms for implementing the rubric more efficiently. With continued refinement, the instrument holds significant potential to support writing instruction and assessment in diverse EFL contexts.

REFERENCES

- Alemdag, E., & Narciss, S. (2025). Promoting formative self-assessment through peer assessment: peer work quality matters for writing performance and internal feedback generation. *International Journal of Educational Technology in Higher Education*, 22(1), 1–26. <https://doi.org/10.1186/S41239-025-00522-4/METRICS>
- Andheska, H., Suparno, S., Dawud, D., & Suyitno, I. (2020). Writing motivation and the ability in writing a research proposal of generation Z students based on cognitive style. *Journal for the Education of Gifted Young Scientists*, 8(1), 87–104. <https://doi.org/10.17478/JEGYS.651436>
- Asli, N. F., Mohd Matore, M. E. E., & Md Yunus, M. (2024). Construct validity of primary trait writing rubrics based on assessment use argument (AUA) validation framework. *Heliyon*, 10(22), e40053. <https://doi.org/10.1016/J.HELIYON.2024.E40053>
- Barkaoui, K. (2010). Variability in ESL Essay Rating Processes: The Role of the Rating Scale and Rater Experience. *Language Assessment Quarterly*, 7(1), 54–74. <https://doi.org/10.1080/15434300903464418>
- Brachman, L. F., & Palmer, A. S. (2010). Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 2010, Vol. 10, No. 2 Book Review. *Language Testing*, 10(2), 41–46.
- Brookhart, S. M. (2013). *How to Create and Use Rubrics for Formative Assessment and Grading*. ASCD.
- Brown, H. D. (2004). *Language assessment: Principle and classroom practices*. Pearson Longman.
- Brown, H. D., & Abeywickrama, P. (2010). *Language Assessment: Principle and Classroom Practices (2nd ed.)*. Pearson Education.
- Chauhan, P. (2022). Fundamentals of Academic Writing: A Literature Review. *Journal of NELTA*, 27(1–2), 161–180. <https://doi.org/10.3126/NELTA.V27I1-2.53201>
- Clarkson, R. (2024). ‘It’s missing the heart of what writing is about’: teachers’ interpretations of writing assessment criteria. *British Educational Research Journal*, 50(1), 134–161. <https://doi.org/10.1002/BERJ.3916>
- Dong, J., Zhao, Y., & Buckingham, L. (2024). Thirty years of writing assessment: A bibliometric analysis of research trends and future directions. *Assessing Writing*, 61, 100862. <https://doi.org/10.1016/J.ASW.2024.100862>
- East, M. (2009). Evaluating the Reliability of a Detailed Analytic Scoring Rubric for Foreign Language Writing. *Assessing Writing*, 14(2), 88–115. <https://doi.org/10.1016/j.asw.2009.04.001>
- East, M., & Slomp, D. (2024). The ethical turn in writing assessment: How far have we come, and where do we still need to go? *Language Teaching*, 57(2), 262–

273. <https://doi.org/10.1017/S0261444823000034>
- Ehsanzadeh, S. J., & Dehnad, A. (2024). Analysis of high-frequency errors and linguistic patterns in EFL medical students' English writing: Insights from a learner corpus. *BMC medical education*, 24(1), 1264. <https://doi.org/10.1186/S12909-024-06242-Z/TABLES/7>
- Elander, J., Harrington, K., Norton, L., Robinson, H., & Reddy, P. (2006). Complex skills and academic writing: A review of evidence about the types of learning required to meet core assessment criteria. *Assessment and Evaluation in Higher Education*, 31(1), 71–90. <https://doi.org/10.1080/02602930500262379;JOURNAL:JOURNAL:CAEH19;PAGE:STRING:ARTICLE/CHAPTER>
- Fatmawati, E., Syathroh, I. L., Siagian, C. B., Anggraini, D. F., & Herman, H. (2024). Understanding the Meaning of an Advertisement Text through Interpersonal Function Analysis. *Anglophile Journal*, 4(1), 30-37. <https://doi.org/10.51278/anglophile.v4i1.875>
- Ferris, D. R., & Hedgcock, J. S. (2023). *Teaching L2 Composition: Purpose, Process, and Practice (4th ed.)*. Routledge.
- Gustafsson, M., & Ganobcsik-Williams, L. (2023). Editorial: The Challenges of Academic Literacy/ies in Teaching Writing: Adaption, Contexts and Conditions. *Journal of Academic Writing*, 13(2), ii–v. <https://doi.org/10.18552/JOAW.V13I2.1063>
- Hyland, K. (2007). Genre Pedagogy: Language, Literacy and L2 Writing Instruction. *Journal of Second Language Writing*, 16(3), 148–164. <https://doi.org/10.1016/j.jslw.2007.07.005>
- Hyland, K. (2013). *Second Language Writing (2nd ed)*. Cambridge University Press.
- Iordanou, K., & Rapanta, C. (2021). “Argue With Me”: A Method for Developing Argument Skills. *Frontiers in Psychology*, 12, 631203. <https://doi.org/10.3389/FPSYG.2021.631203/BIBTEX>
- Jonsson, A., & Svingby, G. (2007). The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Kemdikbud. (2020). *Kurikulum Merdeka Belajar*. Kementerian Pendidikan dan Kebudayaan Indonesia.
- Knoch, U. (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
- Madkur, A. (2024). Multilingual Realities and English Teacher Construction in Indonesian Pesantren: A Narrative Inquiry. *Anglophile Journal*, 4(2), 91-102. <https://doi.org/10.51278/anglophile.v4i2.1044>
- Mukminatien, N., Pratama, R. D., Zagoto, I., Fadilah, A. E., & Amin, R. F. (2025). Developing a Writing Assessment Instrument for English Students on Argumentative Essays. *International Journal of Community Engagement Payungi*, 5(1), 29-42. <https://doi.org/10.58879/ijcep.v5i1.65>
- Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, 9(3), e3292. <https://doi.org/10.1002/REV3.3292>
- 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>
- Olsen, T., & Hunnes, J. (2024). Improving students' learning—the role of formative feedback: experiences from a crash course for business students in academic writing. *Assessment and Evaluation in Higher Education*, 49(2), 129–141. <https://doi.org/10.1080/02602938.2023.2187744>;JOURNAL:JOURNAL:CAEH 19;PAGE:STRING:ARTICLE/CHAPTER
- Pipalova, R. (2024). *Investigating Aspects of Academic Discourse* - Google Books. Charles University, Karolinum Press.
- Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2410101>;WGROU:STRING:PUBLIC ATION
- Rizqi, M., & Faujianor, A. (2024). Validitas Pengembangan Media Pembelajaran Bagan dan Audio pada Materi Wudhu. *Attractive: Innovative Education Journal*, 6(3), 414-422. <https://doi.org/10.51278/aj.v6i3.1702>
- Salihoglu, Y. (2024). The Relationship between Fear of Making Mistakes and Self-Confidence Level in Language Learning: A Review Article. *Linguistic Forum - A Journal of Linguistics*, 6(1), 88–101. <https://doi.org/10.5281/zenodo.14863292>
- Sato, T. (2022). Assessing critical thinking through L2 argumentative essays: an investigation of relevant and salient criteria from raters' perspectives. *Language Testing in Asia*, 12(1), 1–19. <https://doi.org/10.1186/S40468-022-00159-4/TABLES/7>
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge University Press.
- Wubalem, A. Y. (2021). Assessing learning transfer and constraining issues in EAP writing practices. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1), 1–22. <https://doi.org/10.1186/S40862-021-00122-5/TABLES/3>
- Wulandari, B., Rokhmawan, T., & Dewi, A. C. (2023). Initiating an Educational Agricultural Flower Tourism Village through the Construction of Language Landscape and Eco Folklore. *Bulletin of Community Engagement*, 3(2), 161-174. <https://doi.org/10.51278/bce.v3i2.756>
- Yang, R. (2022). An empirical study on the scaffolding Chinese university students' English argumentative writing based on toulmin model. *Heliyon*, 8(12). <https://doi.org/10.1016/J.HELIYON.2022.E12199/ASSET/7A94FB00-36CD-44BF-9D12-4BE189DCD5D9/MAIN.ASSETS/FX4.JPG>
- Yeo, S. W., Signorelli, C., Vo, K., & Smith, G. (2024). A Retrospective Cohort Analysis Comparing Analytic and Holistic Marking Rubrics in Medical Research Education. *Journal of Medical Education and Curricular Development*, 11, 23821205241277336. <https://doi.org/10.1177/23821205241277337>
- Yuskar, B. O. (2024). The Utilization of Three Writing Schemes (TWS) in Teaching Essay Writing to UBSI Students in English Department. *Bulletin of Science Education*, 4(1), 24-34. <https://doi.org/10.51278/bse.v4i1.817>
- Zheng, Y., & Yu, S. (2019). What has been assessed in writing and how? Empirical evidence from Assessing Writing (2000–2018). *Assessing Writing*, 42, 100421. <https://doi.org/10.1016/J.ASW.2019.100421>

Copyright Holder :
© M. Hilmy Hidayatullah et al., (2025).

First Publication Right :
© Bulletin of Science Education

This article is under:
CC BY SA