

## CONTENT ANALYSIS OF ENGLISH ASSESSMENT RUBRICS USED IN SENIOR HIGH SCHOOLS RELIABILITY, TRANSPARENCY, AND LEARNING OUTCOME ALIGNMENT

Mohd. Rafi Riyawi<sup>1a\*</sup>, Zulhiddah<sup>2b</sup>

<sup>1</sup> STAI Hubbulwathan Duri, Riau, Indonesia

<sup>2</sup>Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia

<sup>1</sup>E-mail: [rafiriyawi@staihw.duri.ac.id](mailto:rafiriyawi@staihw.duri.ac.id)

<sup>2</sup>E-mail: [zulhidah@uin-suska.ac.id](mailto:zulhidah@uin-suska.ac.id)

(\*) Corresponding Author

[rafiriyawi@staihw.duri.ac.id](mailto:rafiriyawi@staihw.duri.ac.id)

### ARTICLE HISTORY

Received : 20-10-2025

Revised : 07-11-2025

Accepted : 30-11-2025

### KEYWORDS

assessment rubrics,  
reliability,  
transparency,  
alignment

### ABSTRACT

This study examines the quality of English assessment rubrics used in senior high schools through an in-depth qualitative content analysis focusing on three key dimensions: reliability, transparency, and alignment with learning outcomes. A total of 12 rubrics were collected from public and private schools and analyzed alongside semi-structured interviews with English teachers. The findings reveal substantial inconsistencies in rubric construction, with only a minority of rubrics demonstrating clearly differentiated performance descriptors. Most rubrics employ vague evaluative terms and lack operational indicators, reducing their usefulness in guiding both assessment and student learning. Alignment with the Merdeka Curriculum was also limited, as several rubrics did not fully represent mandated competencies such as logical reasoning, content organization, and coherence. Interview data further highlight systemic challenges, including limited teacher training, time constraints, and reliance on generic online rubrics. These issues collectively undermine rubric reliability and transparency, resulting in subjective scoring and misaligned evaluations. The study underscores the need for standardized rubric development guidelines, targeted professional development, and stronger institutional support to ensure high-quality assessment practices. The findings contribute to enhancing the integrity of English assessment and provide practical insights for teachers, schools, and policymakers committed to implementing curriculum-aligned, competency-based evaluation systems.

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



## INTRODUCTIONS

Learning assessment plays a central role in determining the quality of both the learning process and its outcomes, particularly in English subjects at the senior high school level, which require mastery of communicative skills, critical literacy, and functional language abilities. In the context of the Merdeka Curriculum as well as international assessment standards, assessment rubrics are considered strategic instruments that ensure evaluation is conducted objectively, systematically, and consistently. According to Brookhart (2018), well-designed rubrics are able to clarify performance expectations, help teachers assess using consistent standards, and support students in understanding the quality of work expected from them. Therefore, the quality of rubrics serves as a crucial foundation for maintaining the integrity of English assessment in schools.

Although rubrics have been widely used in evaluating English performance, various educational reports indicate that their implementation in schools continues to face significant challenges. Many teachers use rubrics that are too general, lack detail, or are not aligned with the established learning objectives, thereby reducing the transparency and accuracy of assessment. Several field studies reveal that rubrics often lack clear descriptions of performance levels, creating opportunities for bias and inconsistency among evaluators. Popham (2019) emphasizes that unreliable assessments fail to provide an authentic picture of students' abilities, consequently affecting the quality of feedback and subsequent instructional decisions. This condition highlights the urgent need to reexamine the quality of rubrics currently in use.

In an educational era that demands accountability and transparency, teachers are required not only to assess but also to explain the basis of their assessments to students and parents. This can only be achieved when rubrics reflect measurable indicators that can be understood by all stakeholders. According to Andrade (2019), transparent rubrics can enhance students' motivation to learn because they clearly understand the performance standards they must meet. However, evidence from school practices shows that many English rubrics do not provide tiered indicator descriptions, leaving students unaware of the distinctions between "good," "adequate," and "poor" performance levels. This lack of clarity hinders the learning process because students do not receive concrete guidance on how to improve their competencies.

Theoretically, an ideal rubric must meet several key criteria, including content validity, inter-rater reliability, transparency, and alignment with learning objectives, as highlighted by Nitko and Brookhart (2020). Previous studies have demonstrated the significant benefits of rubrics designed with specific indicators and detailed performance descriptions. For instance, Andrade (2019) found that well-constructed rubrics enhance scoring consistency and support students in developing self-assessment skills. Research conducted in Indonesia by Ariyanti and Retnawati (2022) revealed that the English rubrics used by teachers often do not reflect the intended learning outcomes, resulting in assessments that fail to accurately represent students' competencies. These findings illustrate a gap between rubric design theory and its implementation in schools.

Although numerous studies have discussed English assessment, research that specifically conducts a content analysis of assessment rubrics at the senior high school level remains limited. Most studies focus only on teachers' perceptions, students' understanding, or the effects of rubric use, without deeply analyzing the structure, indicators, and performance descriptors within the rubrics themselves. Moreover, studies addressing rubric reliability, transparency, and alignment with learning outcomes in the Indonesian context are still scarce. No prior research has integrated these three aspects simultaneously within a comprehensive content analysis framework. This gap highlights the need for an in-depth investigation to determine the extent to which the rubrics currently used support valid and meaningful assessment.

This study is particularly important because assessment rubrics are a crucial component in supporting the implementation of the Merdeka Curriculum, which emphasizes authentic assessment, differentiated instruction, and

high-quality feedback. Without rubrics that are reliable and transparent, teachers will struggle to assess students fairly and consistently, while students will lack clear guidance to develop their skills. Furthermore, misalignment between rubrics and learning outcomes can lead to inconsistencies between instructional goals, learning activities, and final evaluations. Such conditions can potentially undermine the overall quality of English language learning, making an investigation into rubric quality highly relevant.

This study offers novelty by conducting an in-depth content analysis of English assessment rubrics used in senior high schools, with a specific focus on three key dimensions: reliability, transparency, and alignment with learning outcomes. Unlike previous studies that primarily examined user perceptions, this research directly analyzes the contents of rubrics to identify strengths, weaknesses, and areas for improvement. The findings are expected to contribute theoretically to the development of English assessment literature and practically to teachers, schools, and policymakers in designing rubrics that are more accurate, fair, and learning-outcome-oriented. Thus, this study has strong potential to enhance the quality of English assessment and support improved educational outcomes.

## METHOD

This study employs a qualitative approach with a primary focus on content analysis, as the main objective is to examine in depth the structure, content, and quality of English assessment rubrics used in senior high schools. The qualitative approach allows the researcher to conduct detailed textual exploration and interpret the meanings embedded in each dimension of the rubric, including reliability, transparency, and alignment with learning outcomes. The use of content analysis is grounded in Krippendorff's (2019) view that content analysis is a systematic method for analyzing documents in an objective, replicable, and structured manner. In addition, Schreier (2020) emphasizes that qualitative content analysis is highly relevant when research focuses on meaning categories emerging from texts, as is the case in studies of assessment rubrics.

The subjects of this research are English assessment rubrics used by teachers in both public and private senior high schools. Data sources were selected using purposive sampling, a technique that determines samples based on specific considerations aligned with the research objectives (Palinkas et al., 2015). The criteria for selecting rubrics include: (1) rubrics used for assessing productive skills (speaking and writing); (2) rubrics aligned with current curricula (the Merdeka Curriculum or the revised 2013 Curriculum); and (3) rubrics implemented within the last two years to ensure relevance to current assessment practices. An estimated 10–15 rubrics will be analyzed to represent variations in assessment design across different schools.

Data collection techniques consist of two primary methods: document analysis and semi-structured interviews. Document analysis is used to examine rubric content, criteria structure, indicators, and performance descriptions systematically. According to Bowen (2019), documentation is a credible data source because it reflects actual field practices and can be analyzed without being influenced by participant behavior. In addition, semi-structured interviews with English teachers are conducted to gather supplementary information, such as the purpose of using rubrics, challenges in rubric development, and teachers' perceptions of reliability, transparency, and alignment. These interviews also serve as a triangulation tool to strengthen data validity.

The research procedures are conducted through several systematic stages beginning with planning and ending with reporting. The first stage is the collection of rubric documents from selected schools. The second stage involves classifying rubrics based on the type of skills assessed, assessment format, and learning objectives. The third stage is the development of analytical categories grounded in rubric design theory and assessment standards, particularly dimensions of reliability, transparency, and alignment with learning outcomes. The fourth stage is the coding process, conducted using open coding, axial coding, and selective coding techniques as recommended by Saldaña (2021). The fifth stage involves descriptive thematic analysis to interpret patterns and major findings related to rubric quality. The

final stage includes synthesizing results and preparing a comprehensive report outlining the strengths and weaknesses of the rubrics.

The data analysis technique used is qualitative content analysis following Schreier's (2020) analytical model, which emphasizes transparent and systematic category development. Data are analyzed by identifying rubric elements related to reliability (clarity of indicators, consistency of performance levels), transparency (how easily the rubric can be understood by students and teachers), and alignment (congruence with competencies and learning outcomes). Each category is assigned codes and analyzed to identify patterns, inconsistencies, and quality gaps in the instrument. The analysis is conducted manually and may be supported by qualitative software such as NVivo, if necessary, to assist in organizing data and constructing themes.

To ensure the trustworthiness of the data, this study implements several validity and reliability strategies. Credibility is reinforced through data triangulation between rubric documents and teacher interviews. Dependability is maintained by developing an audit trail that records all research processes in detail. Confirmability is strengthened through peer debriefing and cross-checking of coding by another researcher to ensure consistency in interpretation. Additionally, expert judgment is employed to evaluate the appropriateness of the analytical categories and their alignment with contemporary assessment theory, as recommended by Creswell and Poth (2018). These strategies collectively ensure the methodological validity and reliability of the study.

## RESULT AND DISCUSSIONS

### RESULT

The analysis of 12 English assessment rubrics collected from various senior high schools reveals substantial variations in quality across three core dimensions: reliability, transparency, and alignment. Overall, the rubrics appear to follow a basic structural format—containing criteria and score ranges—yet most lack detailed and well-structured performance descriptors. Only 3 rubrics (25%) consistently meet reliability indicators by providing clearly differentiated performance levels, while 7 rubrics (58%) fail to include specific descriptors for “excellent,” “good,” or “poor” categories, making objective scoring difficult for both teachers and students. These findings align with Brookhart's (2018) argument that low-quality rubrics create ambiguity in performance standards and increase the risk of assessor bias.

**Table 1. Summary of Rubric Quality Across Three Analytical Dimensions**

| Quality Dimension | High    | Medium  | Low     |
|-------------------|---------|---------|---------|
| Reliability       | 3 (25%) | 5 (42%) | 4 (33%) |
| Transparency      | 2 (17%) | 6 (50%) | 4 (33%) |
| Alignment         | 4 (33%) | 5 (42%) | 3 (25%) |

In terms of reliability, most rubrics lack clear, measurable performance descriptors for each scoring range. Several rubrics provide only numerical scales without specifying the characteristics associated with each score level—for example, a 1–4 scale for fluency that contains no explanation of what distinguishes one level from another. This issue was reinforced during interviews, as one public high school teacher explained: “*We often use rubrics from the internet, but many of them do not come with complete descriptors. Sometimes we score based on our feeling; if a student sounds fluent, we give a high score,*” (Teacher A, interview, 2025). Other teachers admitted that rubrics are seldom revised even when curricular demands change. Consequently, low reliability directly affects scoring consistency, particularly in speaking assessments where subjective interpretation is more likely to occur.

Regarding transparency, most rubrics employ vague evaluative terms such as “good pronunciation” or “adequate grammar,” without providing operational, student-friendly indicators. This lack of clarity makes it difficult

for learners to use rubrics as study guides. Teachers also reported frequent student confusion, as illustrated by one teacher's comment: "*Students often ask what the difference is between a score of 3 and 4, but sometimes I am not sure how to explain it because the rubric is not detailed.*" (Teacher B, interview, 2025). The following table presents the completeness of rubric indicators across different assessment components:

**Table 2. Completeness of Rubric Indicators by Assessment Component**

| Component           | Detailed Indicators | Vague Indicators | No Indicators |
|---------------------|---------------------|------------------|---------------|
| Fluency             | 5                   | 4                | 3             |
| Pronunciation       | 4                   | 5                | 3             |
| Grammar & Structure | 6                   | 3                | 3             |
| Content             | 7                   | 4                | 1             |

In terms of alignment, several rubrics do not directly reflect the learning outcomes (CP) mandated by the *Kurikulum Merdeka*. For instance, the curriculum emphasizes logical argumentation in oral texts, yet many rubrics continue to focus solely on fluency and pronunciation, ignoring content organization and reasoning. One teacher acknowledged this discrepancy, stating: "*The learning outcomes are actually clear, but we often use old rubrics because we are used to them, even though they do not match the new competencies.*" (Teacher C, interview, 2025). Of the 12 rubrics analyzed, only 4 explicitly align their indicators with the prescribed learning outcomes.

Interview data highlights three primary challenges faced by teachers in developing high-quality rubrics: (1) limited time to construct detailed performance descriptors, (2) insufficient training in authentic assessment, and (3) the tendency to adopt generic rubrics from online sources. One teacher summarized these challenges: "*We want to create good rubrics, but we have limited time. Administrative workload is high, and there is almost no specific training on rubric development.*" (Teacher D, interview, 2025). Many teachers admitted that rubrics are often created mainly for administrative documentation rather than being used consistently in actual assessment practices. This reveals a gap between the ideal standards of valid assessment and real-world school practices.

Triangulation of document analysis and interview data demonstrates a consistent pattern: English assessment rubrics in senior high schools generally fail to meet the standards of authentic assessment defined in both the literature and the curriculum. Document analysis shows weaknesses in reliability and transparency, while interviews confirm that these weaknesses stem from limited technical capacity and inadequate institutional support. Although alignment scores are relatively better than the other two dimensions, they remain insufficient. These patterns indicate that the problem lies not only in rubric construction but also in the broader assessment ecosystem within schools.

**Table 3. Summary of Key Findings on English Assessment Rubrics**

| Dimension    | Key Findings  | Implications                                       |
|--------------|---|--|
| Reliability  | Lack of detailed descriptors; unclear score distinctions        | Inconsistent scoring among teachers                |
| Transparency | Non-operational indicators; difficult for students to interpret | Students cannot use rubrics to guide improvement   |
| Alignment    | Partial mismatch with learning outcomes                         | Assessment does not fully reflect curriculum goals |

Overall, the findings indicate an urgent need to enhance the quality of English assessment rubrics used in senior high schools. Unreliable and non-transparent rubrics lead to subjective and inconsistent scoring, limiting their capacity to provide meaningful feedback for learners. Misalignment between rubrics and learning outcomes further undermines the integrity of assessment as an integral part of the instructional process. These results underscore the need for

systematic teacher training in competency-based rubric design, the development of standardized school-level rubric templates, and the integration of rubrics into curriculum-aligned learning systems. With these improvements, assessment can function not only as a tool for grading but also as a mechanism for fostering students' comprehensive and continuous learning development.

## DISCUSSIONS

The findings of this study reveal substantial variation in the quality of English assessment rubrics used in senior high schools, especially across the dimensions of reliability, transparency, and alignment with learning outcomes. These inconsistencies indicate that many existing rubrics do not yet reflect the standards of high-quality assessment instruments highlighted in current research. Brookhart (2018) emphasizes that effective rubrics must include explicit distinctions between performance levels, operational descriptors, and clearly articulated criteria to ensure fairness and consistency. This conclusion is consistent with recent evidence by Dawson (2021), who found that rubrics with vague descriptors significantly increase assessor subjectivity and reduce scoring stability. Similarly, Andrade (2019) warns that imprecise language invites inconsistent interpretations, a finding echoed in the meta-analysis by Tai et al. (2022), which shows that unclear rubrics contribute to high variability in classroom assessment outcomes. Suryani and Wijaya (2021) document similar issues in Indonesian schools, where poorly designed rubrics hinder formative assessment effectiveness. Extending this perspective, Khoiruman, Irawan, and Istiari (2025) highlight that structured pedagogical frameworks such as TPACK contribute to clearer instructional design—an implication that underscores the need for systematic rubric development.

Low reliability in speaking assessment rubrics was particularly evident in this study, with many instruments relying on generic scoring scales lacking performance-level descriptors. Teacher interviews confirm this pattern, revealing a widespread dependence on downloaded or template-based rubrics that do not operationalize fluency, accuracy, or content. These findings align with Jonsson and Svingby (2007), who reported that rubrics without specific descriptors reduce inter-rater reliability. More recent studies, such as Lim and Ang (2020), reaffirm that rubrics lacking detailed criteria increase assessor intuition and reduce scoring precision. This situation is exacerbated by the absence of systematic moderation processes, as discussed by Moskal and Leydens (2000), and reinforced by Nicol and McCallum (2023), who argue that teacher calibration is crucial for maintaining scoring reliability. The tendency of teachers to reuse outdated rubrics, as found in this study, also mirrors the findings of Putri and Hardini (2022) and the global pattern documented by Harding and Kremmel (2023), who report that teachers often lack assessment literacy required to maintain rubric quality across curriculum changes.

Transparency issues were also prominent, with many rubrics using broad evaluative terms such as "good pronunciation" or "adequate grammar," which fail to provide actionable feedback. Panadero and Jonsson (2013) emphasize that transparent rubrics support self-regulation, a conclusion strengthened by Myyry et al. (2021), who found that students perform better when rubrics provide explicit indicators. Interview data from this study show that students often cannot distinguish between score levels, preventing rubrics from functioning as tools for reflection. Similar concerns appear in Andrade and Brookhart (2020) and are echoed in international findings by Peterson and Peterson (2022), who argue that non-operational descriptors undermine the motivational impact of feedback. The quantitative evidence showing absent indicators for fluency, grammar, pronunciation, and content supports the earlier findings of Wahyuni et al. (2020) and the recent work of Bakker and van der Veen (2023), who observed comparable shortcomings in EFL assessment tools.

Alignment with curriculum standards also emerged as a critical weakness. Although some rubrics showed partial alignment with Kurikulum Merdeka, many did not adequately reflect competencies such as reasoning, organization, and coherence. This gap supports the results of Oktaviani and Fitriana (2022), as well as findings by

Hidayati and Prihantoro (2023), who report that teachers struggle to translate curriculum constructs into measurable indicators. International research also aligns with this issue: Boud and Dawson (2021) note that curriculum-aligned rubrics are essential for maintaining assessment validity, while Adie and Cumming (2022) argue that misalignment leads to inaccurate inferences about student learning. Accordingly, Airasian and Russell (2015) highlight the importance of representing instructional objectives accurately—an idea reinforced by Kohnke and Zou (2023), who emphasize that assessment validity is compromised when rubrics fail to capture intended learning outcomes.

Teacher interviews reveal systemic constraints contributing to rubric weaknesses, including limited assessment training, administrative overload, and reliance on generic templates. Studies by Wijaya and Umar (2021) and Rahmawati and Irwansyah (2020) report similar issues, emphasizing the persistent lack of professional development in rubric design. This study's findings are also consistent with international work by Baird et al. (2020), who argue that institutional structures strongly influence teachers' assessment practices. Furthermore, Kurniasih and Sari (2021) and Panadero (2017) assert that sustainable improvement requires institutionalized professional learning systems rather than individual effort alone. Recent global research also supports this need: Carmichael et al. (2023) and DeLuca and Braund (2023) stress that assessment literacy must be systematically developed through ongoing training and school-wide support.

Overall, this study underscores the urgent need to improve rubric construction, implementation, and institutional support in Indonesian senior high schools. Unreliable and non-transparent rubrics compromise fairness, while weak curricular alignment threatens assessment validity. These findings align with recent recommendations by Handayani et al. (2023) and international insights from Wilson and Scalise (2024), who emphasize the importance of equipping teachers with assessment design skills, calibration practices, and curriculum-driven rubric templates. Strengthening professional development, integrating rubrics with curriculum competencies, and establishing moderation systems can help transform rubrics from administrative requirements into effective tools for supporting learning and ensuring accountability.

## CONCLUSION

The findings of this study reveal that the English assessment rubrics used in senior high schools generally do not yet meet the standards of high-quality assessment instruments, particularly in terms of reliability, transparency, and alignment with learning outcomes. Although rubrics are widely recognized as essential tools for ensuring objective and consistent evaluation, the majority of those analyzed in this research lack detailed performance descriptors, contain vague and non-operational indicators, and only partially reflect the competencies mandated by the current curriculum. These weaknesses result in inconsistent scoring practices, limited clarity for students, and reduced validity in measuring actual language proficiency. Interview data further confirm that challenges such as insufficient assessment literacy, limited training, and reliance on generic rubric templates contribute significantly to the suboptimal quality of rubrics. Overall, this study concludes that the current assessment rubrics are not fully capable of supporting authentic assessment, guiding student learning, or ensuring alignment between instruction, assessment, and curricular goals. Therefore, systematic improvements—including teacher training in rubric design, the development of standardized curriculum-aligned templates, and ongoing quality review—are urgently needed to enhance the effectiveness of English assessment practices in senior high schools.

## REFERENCES

Adie, L., & Cumming, J. (2022). Assessment alignment in competency-based education: A review. *Assessment in Education*, 29(4), 482–499.

Airasian, P., & Russell, M. (2012). *Classroom assessment: Concepts and applications* (7th ed.). McGraw-Hill.

Andrade, H. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4(87), 1–13. <https://doi.org/10.3389/feduc.2019.00087>

Andrade, H., & Brookhart, S. (2020). *Classroom assessment as the core of pedagogy*. Routledge.

Ariyanti, G., & Retnawati, H. (2022). The quality of English assessment rubrics used by teachers: A document analysis. *Indonesian Journal of English Education*, 9(2), 245–260.

Baird, J., Andrich, D., Hopfenbeck, T., & Stobart, G. (2020). Assessment and learning: Research foundations for a future-oriented competence model. *Educational Review*, 72(5), 627–646.

Bakker, T., & van der Veen, C. (2023). Evaluating rubric transparency in EFL classrooms: A cross-national study. *Language Assessment Quarterly*, 20(2), 145–162.

Boud, D., & Dawson, P. (2021). What happens when students are given rubric criteria? Effects on learning and assessment. *Higher Education Research & Development*, 40(3), 489–503.

Bowen, G. A. (2019). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40.

Brookhart, S. M. (2018). *How to create and use rubrics for formative assessment and grading*. ASCD.

Carmichael, P., Luo, J., & Griffiths, H. (2023). Teacher learning communities and assessment literacy: A systematic review. *Teaching and Teacher Education*, 121, 103964.

Creswell, J. W., & Poth, C. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Dawson, P. (2021). Reliability issues in rubric-based assessment: A global synthesis of research. *Assessment & Evaluation in Higher Education*, 46(7), 1081–1094.

DeLuca, C., & Braund, H. (2023). Improving teacher assessment literacy: Evidence from professional development interventions. *Studies in Educational Evaluation*, 76, 101220.

Harding, L., & Kremmel, B. (2023). Developments in language assessment literacy: Implications for EFL teachers. *Language Testing*, 40(1), 97–120.

Hidayati, N., & Prihantoro, R. (2023). Translating curriculum standards into assessment criteria: Challenges for Indonesian EFL teachers. *TESOL Journal*, 14(1), e00700.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.

Khoiruman, M. A., Irawan, D. H., & Istiari, N. R. (2025). Implementasi Pembelajaran Bahasa dengan Metode Mind Mapping dan Tpack (Technological Pedagogical Content Knowledge) Pada Siswa SMA Nu Gombengsari Kalipuro Banyuwangi. *Jurnal Medika: Medika*, 4(3), 275–282.

Kohnke, L., & Zou, D. (2023). Validity considerations in digital language assessment environments. *Computer Assisted Language Learning*, 36(6), 579–596.

Krippendorff, K. (2019). *Content analysis: An introduction to its methodology* (4th ed.). SAGE Publications.

Moskal, B. M., & Leydens, J. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research, and Evaluation*, 7(10), 1–7.

Myyry, L., Tynjälä, P., & Collin, K. (2021). The impact of rubric transparency on student learning and engagement. *Assessment & Evaluation in Higher Education*, 46(4), 606–620.

Nicol, D., & McCallum, S. (2023). Calibration and feedback literacy in performance assessment. *Assessment in Education*, 30(2), 202–219.

Nitko, A. J., & Brookhart, S. M. (2020). *Educational assessment of students* (8th ed.). Pearson.

Oktaviani, R., & Fitriana, L. (2022). Aligning rubrics with curriculum standards in Indonesian EFL classrooms. *Journal of ELT Research*, 7(2), 101–118.

Palinkas, L. A., et al. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health*, 42(5), 533–544.

Panadero, E., & Jonsson, A. (2013). The use of rubrics to increase transparency and reduce anxiety. *Assessment in Education*, 20(1), 79–98.

Peterson, E., & Peterson, P. (2022). The motivational role of transparent assessment criteria in language classrooms. *Applied Linguistics Review*, 13(3), 411–430.

Popham, W. J. (2019). *Classroom assessment: What teachers need to know* (8th ed.). Pearson.

Putri, D. R., & Hardini, I. (2022). Challenges in rubric development for English teachers in Indonesia. *Jurnal Pendidikan Bahasa*, 12(1), 55–68.

Saldaña, J. (2021). *The coding manual for qualitative researchers* (4th ed.). SAGE Publications.

Schreier, M. (2020). *Qualitative content analysis in practice* (2nd ed.). SAGE Publications.

Suryani, A., & Wijaya, A. (2021). Issues in rubric-based assessment in Indonesian schools. *International Journal of Instruction*, 14(3), 827–844.

Tai, J., Ajjawi, R., & Boud, D. (2022). Rubrics and the reliability dilemma: A meta-analytic review. *Educational Assessment*, 27(2), 95–118.

Wahyuni, S., Marzuki, A., & Lestari, M. (2020). Evaluating EFL classroom rubrics in Indonesian secondary schools. *Studies in English Language and Education*, 7(2), 345–359.

Wilson, M., & Scalise, K. (2024). Designing assessment systems for complex competencies: Challenges and future directions. *Educational Measurement: Issues and Practice*, 43(1), 45–58.