



Contents lists available at ojs.aeducia.org

Indonesian Journal of Innovative Teaching and Learning

Volume 2, Issue 1 (2025), 10.64420/ijitl.v2i1

Journal homepage: <https://ojs.aeducia.org/index.php/ijitl>

IJITL

E-ISSN 3048-3816

P-ISSN 3062-9772

Review Article

Read Online: <https://doi.org/10.64420/ijitl.v2i1.197>

Open Access

A Review of Rubrics in Education: Potential and Challenges

Jen Hua Ling

University of Technology Sarawak, Malaysia

ARTICLE HISTORY

Received: January 03, 2025

Revised: January 27, 2025

Accepted: February 16, 2025

Published: February 27, 2025

CONTENT

[Introduction](#)

[Method](#)

[Result and Discussion](#)

[Implications and Contributions](#)

[Limitation & Future Research Directions](#)

[Conclusion](#)

[Acknowledgments](#)

[Author Contribution Statement](#)

[Conflict of Interest Statement](#)

[Ethical Approval Statement](#)

[References](#)

[Article Information](#)

ABSTRACT

Background: Rubrics have been praised for offering clear criteria, enhancing objectivity, and promoting consistency in assessment. However, challenges remain regarding their application in diverse educational environments.

Objective: In recent years, the use of rubrics in educational assessment has gained significant attention due to their potential to enhance transparency, consistency, and learning outcomes. However, debates persist regarding their practical implementation, effectiveness, and limitations in various educational contexts.

Objective: This study reviews the literature on rubrics, focusing on their role in learning facilitation, fair assessment, reliability, and validity. **Method:** The research follows a literature study approach, analyzing data from journals, scientific articles, and literature reviews. The analysis process includes data collection, classification, presentation, and conclusion drawing, with data triangulation used for validation. **Results:** Teachers should have the flexibility to decide whether to use rubrics. If adopted, rubrics must be well-designed and implemented by incorporating essential characteristics, avoiding common pitfalls, and continuously improving them to ensure fairness, reliability, and validity.

Conclusion: Rubrics can be effective assessment tools when properly designed and implemented, but they require ongoing refinement to maximize their benefits.

Contribution: This study provides insights into best practices for rubric use, highlighting key factors for effective implementation in educational settings.

KEYWORDS

Rubric; Facilitate learning; Fair assessment; Validity; Reliability

1. INTRODUCTION

A rubric is a scoring tool for qualitative ratings of authentic or complex student work (Jonsson & Svingby, 2007). Teachers use it to communicate expectations for an assignment (Andrade et al., 2008). It can be created for various subjects (Tierney & Simon, 2004), particularly when the constructed response is fairly significant (Popham, 1997). Examples include group activities, extended projects, oral presentations, literature reviews, critical thinking, and writing (Moskal, 2000; Reddy & Andrade, 2010). As a common assessment tool, rubrics are designed to facilitate learning and ensure fair evaluation. However, conflicting perspectives exist regarding their benefits and limitations, leading to confusion among educators about properly implementing rubrics in the classroom.

* **Corresponding Author:** Jen Hua Ling, lingjenhua@uts.edu.my

School of Engineering and Technology, University of Technology Sarawak, Malaysia

Address: School of Engineering and Technology, University of Technology Sarawak, 96000 Sibu, Sarawak, Malaysia

How to Cite (APA 7th Edition):

Ling, J. H. (2025). A Review of Rubrics in Education: Potential and Challenges. *Indonesian Journal of Innovative Teaching and Learning*, 2(1), 1-14. <https://doi.org/10.64420/ijitl.v2i1.197>



Copyright © 2025 by the author(s). This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) License (<https://creativecommons.org/licenses/by-sa/4.0/>)

A rubric comprises evaluation criteria, quality definitions, and a scoring strategy (Popham, 1997). Evaluation criteria outline the content of an assessment (Parke, 2001), which is considered while determining the quality of student work (Reddy & Andrade, 2010). Quality definitions describe the differences in students' responses (Popham, 1997), distinguishing levels of quality from excellent to poor (Andrade et al., 2008). Scoring strategies involve the use of a scale for assessment. Some studies suggest that rubrics enhance clarity in expectations and transparency in assessment, help students understand their learning targets, and enable them to monitor and evaluate their progress. However, criticisms have also emerged, arguing that rubrics may restrict students' creativity and create an overly structured environment, thereby limiting interpretation and free expression. Additionally, poorly designed or implemented rubrics may fail to meet expectations, particularly for less experienced teachers.

Rubrics contribute to assessment reliability, validity, and transparency (Jeong, 2015). In a formative assessment, rubrics can both teach and evaluate (Reddy & Andrade, 2010). However, concerns exist that rubrics may pose more problems than solutions (Li & Lindsey, 2015). Some educators argue that rubrics have not improved grading reliability (Rezaei & Lovorn, 2010). Rubrics do not necessarily facilitate valid judgements (Bryant et al., 2016). Most rubrics have been misrepresented as instructional aids when, in reality, they have no educational impact at all (Popham, 1997). These conflicting perspectives contribute to teacher perplexity regarding the effective use of rubrics in the classroom. The benefits of rubrics may have been somewhat overstated. The positive effects of rubrics are widely accepted despite criticism. However, not all critiques lack merit. Criticism can be approached more open-mindedly. Understanding the limitations allows for the judicious and sparing use of rubrics.

Rubrics as an assessment tool have become a common practice in education. Previous research indicates that rubrics improve clarity in expectations, help students understand assessment criteria, and promote fair and transparent evaluation. Furthermore, rubrics are believed to enhance student engagement in the learning process and facilitate more effective self-assessment. However, some studies also suggest that rubrics limit students' creativity and lead them toward an overly structured learning approach. Moreover, inadequately designed and implemented rubrics can reduce effectiveness and create confusion among educators and students.

Despite numerous studies discussing the benefits and challenges of using rubrics, comprehensive literature does not examine how rubrics can be optimized for various learning contexts. Another identified gap is the lack of practical guidelines for educators in designing and implementing effective rubrics and limited research exploring the long-term impact of rubrics on student learning outcomes. Additionally, there is a need to understand how rubrics can be adapted to different disciplines and educational levels while ensuring that they do not hinder creativity and flexibility in learning.

This study contributes new insights by critically reviewing existing literature, identifying the potential and challenges of using rubrics, and offering practical recommendations for educators. This comprehensive approach includes an analysis of how rubrics can be designed to facilitate deep learning, ensure fair assessment, and enhance the reliability and validity of evaluations. Furthermore, this study highlights the importance of training educators in developing and implementing rubrics and the need for further research to explore rubric adaptations in various educational contexts. Thus, this study not only addresses gaps in the literature but also provides practical guidance for improving the effectiveness of rubrics in education. This paper reviews literature debates on rubrics, exploring learning facilitation, fair assessment, reliability, and validity. It identifies the rationale behind these arguments, aiming to present a balanced perspective for effective use in classes. The aim is to present a balanced perspective on the appropriate use of rubrics. A rubric has potential and limitations.

2. METHOD

2.1 Research Design

The study used a systematic literature review design to analyze the use of rubrics in education. The main objective of this study was to present a balanced perspective on the potential and challenges of using rubrics, as well as provide practical recommendations for educators. Literature study is a series of activities related to library data collection methods, reading and recording, and processing research materials. The researcher used this method to review the literature debate on rubrics, exploring learning facilitation, fair assessment, reliability, and validity. Literature study is used as a required activity in research, especially academic research whose main purpose is to develop theoretical aspects and practical benefits. Through this desk study, the researcher had a broader and deeper dive into the literature debate on rubrics, exploring learning facilitation, fair assessment, reliability, and validity for effective use in the classroom. With this approach, this research seeks to make a comprehensive contribution to the understanding and practice of using rubrics in education and offer solutions to overcome the challenges faced in their implementation.

2.2 Data Collection

The data used came from journals, scientific articles, and literature reviews on rubrics, exploring learning facilitation, fair assessment, reliability, and validity.

2.3 Data analysis

After the data is collected, the data is then analyzed. The steps in conducting data analysis are as follows (1) data collection; (2) data reduction or data classification; (3) data presentation or data analysis; (4) the final process of drawing conclusions.

2.4 Data Triangulation

Data validity checking uses data triangulation. Researchers use source triangulation to test the credibility of data by checking the data that has been obtained from various data sources such as the results of data collection obtained from journals, scientific articles, and literature reviews.

3. RESULT AND DISCUSSION

3.1 Result

a) Learning Facilitation

Rubrics were claimed to have the potential to promote learning and improve instruction due to their explicit expectations and criteria (Jonsson & Svingby, 2007). Evaluation criteria allowed students to apply critical judgements to their work, resulting in significantly higher rates (Abram et al., 2002). Petkov & Petkova (2006) found higher project grades among students using rubrics than those without, with the mean percentage grade increasing from 95.125% to 96%. However, this minor increase is insufficient to prove the rubrics' effectiveness in promoting learning.

Rubrics do not always guarantee quality work. In Andrade's (2001) study, students with rubrics received higher scores on only one of the three essays. Covill (2012) discovered that students given a rubric did not write better than those without any. Green & Bowser (2006) observed no significant differences in scores between students with and without rubrics. Kavanagh & Luxton-Reilly (2016) even reported that students trained to use rubrics performed no better than their untrained counterparts.

Students generally prefer rubrics over other techniques, particularly unclear subjective grading scales (Bolton, 2006). They value rubrics for clarifying the targets for their work, allowing them to regulate their progress, and making grades transparent and fair (Van Helvoort, 2012). Despite not being strictly proven, students perceive that they learn better with rubrics (Jonsson & Svingby, 2007). Van Helvoort's (2012) students used rubrics in their work and tended to have the criteria "in mind" for their assessments in subsequent courses. Rubrics enable students to identify critical issues, reduce their efforts on unnecessary aspects, reduce uncertainty and ambiguity, and provide feedback on weaknesses (Bolton, 2006). These benefits are more apparent among adults (Bolton, 2006). Students occasionally express dissatisfaction with rubrics, particularly for limiting creativity and creating a restrictive environment with little room for interpretation (Bolton, 2006). They also complain about having to meet unusual or unfathomable expectations (Andrade & Du, 2005). Some students even make no distinction between acceptable and unacceptable expectations, instead just giving their teachers what they want (Andrade & Du, 2005). This attitude defeats the purpose of rubrics in facilitating learning. Students agree that rubrics help them focus their efforts, produce quality work, earn better grades, and feel less anxious (Andrade & Du, 2005). However, Panadero & Romero (2014) noticed students using rubrics exhibited more actions detrimental to learning and higher stress levels. Rubrics may not reduce student anxiety for non-cognitive assessments, such as oral presentations, given the common fear of public speaking among the general population (Galván-Sánchez et al., 2017).

Having the right attitude toward rubrics is crucial. Simply distributing rubrics cannot be expected to influence student work (Reddy & Andrade, 2010). Students must be taught to actively use rubrics for self- and peer assessments (Reddy & Andrade, 2010). Without adequate opportunity to learn using criteria, there is little evidence that students use them (Cohen et al., 2002).

Rubrics outline the assessment criteria. However, gaps exist between teacher expectations and student interpretations (Lea & Street, 1998). Students who attend to a rubric do not always effectively address the specified criteria (Andrade, 2001). Some attempts to meet the requirements were limited or clunky (Andrade, 2001). Li & Lindsey (2015) observed that students had difficulties recognising the focal points converged by teachers in a rubric; teachers converged on a relatively small set of focal points while the students held diverse views; and even when the

same set of focal points was chosen, teachers and students ranked them differently. Rubrics are not entirely self-explanatory (Andrade, 2005). Disparities in interpreting rubrics are more prominent among students than teachers (Li & Lindsey, 2015). Relative to teachers, students have a less developed sense of the criteria for judging their work (Boud, 1989). Due to inadequate reading skills, students often interpret rubrics by individual words rather than entire phrases (Li & Lindsey, 2015). Therefore, unless teachers train students to use evaluation criteria and provide feedback using evaluation criteria, the favourable effects of rubrics are unlikely (Cohen et al., 2002).

The choice of wording in rubrics is critical (Tierney & Simon, 2004). Rubrics are often denounced for subjective language use in the criteria (Sundeen, 2014). Any rubric with abstract, vague, or ambiguous language is likely ineffective (Li & Lindsey, 2015). For example, words like "sophisticated," "adequate," and "simple" can generate vastly different representations in different students (Li & Lindsey, 2015). Students might not recognise quality work even if they know what to look for (Orsmond et al., 1996). Addressing vague and ambiguous language is perhaps the most challenging aspect of rubric design (Li & Lindsey, 2015). Nevertheless, using precise language does not always improve the effectiveness of rubrics (Li & Lindsey, 2015). This may result in overly elaborated explanations, overwhelming students with lengthy and sophisticated rubrics. Most rubrics nowadays are lengthy and laden with details (Popham, 1997). Detailed rubrics may clarify how to ascertain the quality of a student's response and increase between-rater agreement, but excessively detailed rubrics turn teachers off (Popham, 1997).

Precise language bridges the gap between what a rubric states and how a student interprets it. It minimises subjectivity, narrows interpretation, and encourages a formulaic approach when using rubrics. This raises concerns about restricting students' problem-solving, decision-making, and creativity (Chapman & Inman, 2009). Falchikov (1986) recommended being less rigid and allowing credit for creativity and originality. Depending on assessment goals, a balanced mix of precise language and subjectivity might work.

Many rubrics have problems with contextual consistency. There are basic consistency, where the performance criterion changes from level to level; and negative-positive consistency, marked by a dichotomous (negative or positive) tone (Tierney & Simon, 2004). Inconsistent rubrics are less effective as learning tools. Students can still learn, but not necessarily the intended learning goal (Tierney & Simon, 2004). Rubrics should seek a balance between generalised wording and detailed description; be concrete and concise; avoid jargon; use parallel language; differentiate performance levels; and avoid describing low levels of student performance in negative terms (Tierney & Simon, 2004).

Furthermore, teachers may recognise good performance but struggle to articulate exactly their expectations (Higgins et al., 2002). This challenge arises because the conceptions of quality usually take the form of tacit knowledge (Higgins et al., 2002). Qualitative judgements involve multiple criteria, which are "fuzzy" (Sadler, 1989). Of many potential criteria, only a small subset is typically used at any one time, and a fixed set of criteria is not uniformly applicable to all students in an assessment (Sadler, 1989). This complexity makes it challenging for teachers to specify all the relevant criteria in advance (Sadler, 1989). This underscores the importance of calibrating rubrics for improved assessment, which is currently underemphasized.

Rubrics can be used as part of a student-centred approach to assessment (Reddy & Andrade, 2010). They can encourage learning, but they are not always effective. Simply constructing and distributing written grade descriptors provides little or no benefit to students' learning (Hendry et al., 2012). For example, comments on a rubric do not help students write better, as even the most carefully chosen complementary comments may not create meaningful feedback about students' work (Wilson, 2007). The key lies in facilitating the learning process. For criteria challenging to students, teachers may supplement instruction as necessary (Cohen et al., 2002). Teachers who use rubrics to monitor the development of advanced thinking skills and give feedback have more positive attitudes towards rubrics (Kutlu et al., 2010). Teachers also need to know their students well, understand their past learning history and beliefs, diagnose their difficulties, induce from learner work common patterns to prioritise teaching points, probe learners to think and elaborate, and make real-time decisions to provide useful guidance (Huang, 2012).

Rubrics are not a replacement for good instruction (Andrade, 2005). Even a fabulous rubric does not change the fact that students need models, feedback, and opportunities to ask questions, think, and revise (Andrade, 2005). In principle, clearly articulated and easily accessible criteria can improve student performance by empowering them to evaluate their work (Abram et al., 2002). Evaluation criteria motivate students to be more self-critical and increase their efforts for better performance (Cohen et al., 2002). By having evaluations more frequently, specifically, and clearly, students become more productive (Abram et al., 2002). Peer and self-assessments help students think more, work in a more structured way, and be more critical (Falchikov, 1986; Orsmond et al., 1996). Properly designed rubrics can effectively convey those features and stimulate learning. Using rubrics for self-evaluation renders valid assessments of students' own performance, whereas using rubrics in peer assessment provides an effective learning

strategy (Hafner & Hafner, 2003). However, caution is advised when employing self- and peer assessments. Self-assessment can be comparable to traditional tutor assessment, but peer assessment corresponds less well with the tutor- or self-grading (Falchikov, 1986). Students are not always good at peer and self-assessments, even with a rubric in hand (Andrade, 2005). Students found them hard and time-consuming (Falchikov, 1986). Self-assessments can be misguided or delusional, and peer assessments can be cruel or disorienting (Andrade, 2005). Without proper supervision, self- and peer-assessments using rubrics may fall short of expectations.

Teachers may co-create rubrics with students (Andrade & Du, 2005). This can promote learning, make the assessment goals and qualities transparent, and have students use rubrics to guide peer and self-assessment (Reddy & Andrade, 2010). The process includes discussing examples with students, brainstorming criteria, and using the resulting list to draft comments (Andrade, 2005). Co-creating rubrics with students improved their writing quality and attitudes towards writing (Johnson & Gelfand, 2013). The outcome, however, is not always positive. The process is time-consuming. The motivated students may be anguished over what to do with it, while the unmotivated ones simply record it and promptly forget about it (Andrade, 2005).

b) Fair Assessment

In an assessment, teachers may use different criteria to evaluate students' work, have different understandings of shared criteria, and have different senses of appropriate standards (Bloxham et al., 2016). Without rubrics, an evaluation would be more subjective (Chan & Ho, 2019). Rubrics ensure consistent marking, especially when multiple teachers are involved (Chan & Ho, 2019). They standardise evaluation methods (Chan & Ho, 2019), reducing possible rater bias (Galván-Sánchez et al., 2017). Upon grading disagreements, teachers can discuss their differences based on the rubrics (Chan & Ho, 2019). Rubrics promote transparency and fairness in evaluation (Chan & Ho, 2019). With rubrics, students have clear directions instead of feeling confused by different requirements from different teachers (Chan & Ho, 2019). Rubrics help clarify assessment criteria and orient students to the desired performance levels (Galván-Sánchez et al., 2017). They help students identify their strengths and weaknesses when used for feedback, knowing "what counts" made grades seem fair (Andrade & Du, 2005). Students view rubrics as a fair way to grade, justifying their grades by telling what they did wrong, what they need, and what they are lacking (Andrade & Du, 2005). Sometimes, students perceive that rubrics involve components that teachers consider important but lack aspects relevant to them, such as effort (Panadero & Romero, 2014). Students define fairness as a valid measure of what they deem to be meaningful and worthwhile learning (Sambell et al., 1997).

Students think teachers should not mark assignments based on their impressions of a student's personality or non-academic performance, rubrics ensure teachers mark objectively and minimise potential bias towards particular students (Chan & Ho, 2019). They help minimise variables, including teachers' mood, that can affect marking (Chan & Ho, 2019). Additionally, rubrics provide a defensible platform for grievances if students protest their grades (Adedoyin, 2013). Rubrics cannot effectively improve perceived fairness without proper facilitation and learning processes. According to Pepper & Pathak (2008), explicitness of grading criteria, frequency of feedback, and proactive instruction techniques increase the perceived fairness. A consistent assessment cannot be achieved without the awareness and commitment of teachers. Many teachers use rubrics without considering their reliability (Rezaei & Lovorn, 2010). They may not understand the outcomes they are supposed to judge, may not agree with, ignore, or choose not to adopt the criteria, and may interpret the criteria differently (Bloxham et al., 2016).

Additionally, teachers may not completely adhere to rubrics when grading. Some teachers give points for citations in essays that have none, they grade based on their overall impression rather than following rubrics (Rezaei & Lovorn, 2010). Teachers are also strongly influenced by the trivial mechanics and superficial aspects of assessments, students who write neatly and display better basic writing mechanics regularly receive higher marks (Rezaei & Lovorn, 2010). Teachers sometimes allocate grades for criteria that are not attempted (Kavanagh & Luxton-Reilly, 2016), such as penalising students for late assignments, offering bonus marks for extra work, and restricting the time available for students to complete assessment tasks (Scott et al., 2014). Such inappropriate practices should be avoided.

Bias grading can be conscious or unconscious influenced by extraneous factors such as student attractiveness and performance in prior courses (Malouff, 2008). Teachers might assign higher grades because a student is very nice, has done well in other work, seems hardworking and conscientious, seems very interested in the class, seems eager for a high grade, needs a high grade for some important purpose, has a certain gender or ethnic status, is attractive, looks, speaks, or acts like a very intelligent person, or is connected to the instructor. Thus, Malouff (2008) recommended keeping students anonymous during grading.

Feedback provided by a rubric surpasses a mere grade. However, untrained teachers may simply use rubrics to justify biased assessments (Rezaei & Lovorn, 2010). Insufficient knowledge of rubrics, teachers fail to use them

appropriately and fall short of evaluating students correctly (Kutlu et al., 2010). This leads to the impression that rubrics make grading quicker and more responsive, but not better and fairer grades (Reynolds-Keefer, 2010).

While a consensus on fair assessment may be lacking, perceived fairness can be improved by transparent assessment criteria from the outset; objective evaluation of all students using the same criteria; justification and relevance of criteria; clear and easily understood quality definitions; reasonable expectations for students; and consistent calculation of grades using rubric-defined scoring strategies. These principles embody transparency, objectivity, validity, explicitness, equality, and reliability.

Murillo & Hidalgo (2017) categorized fair assessment into egalitarian and equitable conceptions. Egalitarianism prioritises objectivity, reliability, transparency, and validity, which are associated with legal or general justice. Equitable emphasises adaptation, diversification of tests, qualitative assessment, and consideration of students' efforts and attitudes, which resemble distributive or social justice (Murillo & Hidalgo, 2017).

Primary and secondary school students have different positions on fair assessments. The former perceives it as closer to social justice and to the importance that education contributes to compensating inequalities, whereas the latter considers it objective and equal for all (Murillo & Hidalgo, 2017).

Teachers may consider aligning rubrics with students' perceived fairness without compromising assessment goals. When evaluations are seen as sound, students will direct their efforts towards improving the evaluations they receive (Abram et al., 2002; Cohen et al., 2002). Conversely, if students consider assessments unfair, the influence on their subsequent learning may be considerably lower (Murillo & Hidalgo, 2017).

Scott et al. (2014) outlined five key principles of fair assessment: Understanding and addressing the impact of assessment on individual students and their families; Differentiating assessment to accommodate each student's ability and socio-cultural background; Challenging indefensible and illogical assessment practices; Avoiding overwhelming assessments in terms of frequency, intensity, and intrusiveness; Not using assessment to counter inappropriate student behaviour or reward desired behaviour. These principles may be applied to rubrics.

c) Reliability

Reliability, i.e., the consistency of assessment scores (Moskal & Leydens, 2000), ensures that students are graded consistently (Kavanagh & Luxton-Reilly, 2016). It is crucial for valid subjective assessments. Without reliability, subsequent analyses of teachers' ratings may yield spurious results (Stemler, 2004), leading to poor instructional judgements.

There are inter-rater reliability and intra-rater reliability (Kavanagh & Luxton-Reilly, 2016; Moskal & Leydens, 2000). Inter-rater reliability is the consistency of scores assigned by two independent raters (Reddy & Andrade, 2010). Different raters may have different criteria and judgements about quality work. Without rubrics, two raters may not assign the same score to a response. A rubric might not eliminate variations between raters, but a well-designed rubric can reduce discrepancies (Moskal & Leydens, 2000).

Intra-rater reliability concerns the consistency of scores given by the same rater at different times (Reddy & Andrade, 2010). This involves factors irrelevant to the assessment, like the teacher's mood, condition, and perception of a student. A teacher gets tired over time, and thus, certain responses, if rated earlier, may receive different scores (Moskal & Leydens, 2000). Also, a failing student's correct response may be more critically analysed than that of a good student. The inconsistent scores are due to the teacher rather than true differences in student performances (Moskal & Leydens, 2000).

Reliability can be further classified into inter-rater consensus, inter-rater consistency, intra-rater consensus, and intra-rater consistency (Panadero et al., 2013). Inter-rater consensus involves different assessors awarding the same score for the same performance, while inter-rater consistency relates to the score distribution patterns across a set of assessors. Intra-rater consensus measures the reliability of an assessor grading an assessee on two different occasions, whereas intra-rater consistency gauges the reliability of an assessor grading different assessees on a single occasion (Panadero et al., 2013).

Students often feel their scores are influenced by teachers' subjective judgements. Ideally, the same score should be assigned, regardless of when a student completes the assessment, when the response is rated, and who rated the response (Moskal & Leydens, 2000). However, teachers' judgements are prone to bias and error, potentially compromising the rating quality (Knoch et al., 2007).

Rubrics improve reliability by adding restrictions to assessments (Jonsson & Svingby, 2007), establishing a relatively common interpretation of student performance (Reddy & Andrade, 2010). Hafner & Hafner (2003) observed remarkably uniform rating behaviour between teachers and students in a peer assessment involving rubrics. Also, Simon & Forgette-Giroux (2001) reported a 75% agreement between teacher and student ratings.

However, rubrics do not always give reliable ratings. Rezaei & Lovorn (2010) noticed that rubrics significantly increased grade variance. Li & Lindsey (2015) noted score variations both among students and teachers; student scores varied more than those of teachers; and teacher scores varied significantly with more complex performance descriptions. When more criteria are involved, it is harder to converge on a core group of criteria (Li & Lindsey, 2015).

Rubrics in themselves might not offer sufficient reliability for summative assessments (Jonsson & Svingby, 2007). Factors contributing to inconsistency include differing perceptions of quality, student characteristics, rater bias, and vague or ambiguous language (Li & Lindsey, 2015). Other factors involve rater characteristics towards severity or leniency, rater training experience, rater's language background, and task variability (Jeong, 2015).

Reliability can be improved by adding assessment restrictions, using benchmarks, training raters, and using different scoring methods (Jonsson & Svingby, 2007). Restrictions are added by increasing the specificity of rubrics. This improves inter-rater reliability but leads to reductivity (Kavanagh & Luxton-Reilly, 2016). Overly reducing the breadth of assessment reduces students' efforts to that which meets the criteria, rubrics simplify complex evaluation processes but trade richness, complexity, and difference for efficiency, consistency, and convenience (Bennett, 2016).

Benchmarks, or exemplars, can be written descriptions or actual work samples. They help clarify rubrics. However, benchmarks should be carefully chosen since scoring depends heavily on them (Jonsson & Svingby, 2007). If a benchmark mismatches a rubric, it may override the rubric. To avoid this, multiple examples demonstrating various approaches to the same task may be provided (Jonsson & Svingby, 2007).

Teachers can enhance student performance by explaining exemplars (Hendry et al., 2012). Providing insights outperformed pointing out errors or presenting exemplars without discussion. The students who were told why exemplars were graded the way they were outperformed those informed about the errors in the exemplars, as well as those given exemplars without discussion. However, students may use exemplars more than rubrics and imitate the exemplars. Case- or scenario-based assignments may prevent students from directly copying (Hendry et al., 2012).

Teachers should be trained to effectively design and employ rubrics (Rezaei & Lovorn, 2010). They should also master the subject, know the focus of grading, and calibrate rubrics with pilot samples before actual grading (Zhang et al., 2015). Experienced, well-trained teachers do not rate based on their overall impressions but follow rubrics (Jeong, 2015).

Teachers should be aware of rater effects influencing assessment reliability. These include the severity effect, where raters consistently rate too harshly or too leniently; the halo effect, where raters rate based on their overall impression; the central tendency effect, where raters rate at or near the scale midpoint, avoiding extremes; inconsistency, where raters apply multiple rating scale categories inconsistently and display more random variation; and the bias effect, where raters rate unusually harshly or leniently about one aspect (Knoch et al., 2007).

Training can eliminate extreme differences, increase raters' self-consistency, and reduce individual biases (Knoch et al., 2007). However, training might never completely eliminate differences (Jonsson & Svingby, 2007), and the positive effects may not last (Knoch et al., 2007). Lumley & McNamara (1995) observed different rater behaviour in a 1-month gap.

More quality levels in rubrics can improve inter-rater reliability but not consensus agreement. High consensus agreement can be attained with fewer quality levels (Jonsson & Svingby, 2007). Teachers can rate more accurately on a two-level scale than on a four-level scale. The former requires minimal training, whereas the latter requires experience and training (Williams & Rink, 2003).

It is common to moderate rater effects through double- or triple-ratings (Knoch et al., 2007). Two raters enhance reliability (Baker et al., 1996), enough for an acceptable inter-rater agreement under restrained conditions (Jonsson & Svingby, 2007). However, this may not be cost-effective, doubling the expenses for only a slight increase in reliability (Klein et al., 1998).

This raises an overhead issue since grading is paid work in the education sector (Kavanagh & Luxton-Reilly, 2016). The time and expense involved are not trivial (Dunbar et al., 2006). Zhang et al. (2015) reported that holistic scoring of 300 samples took 1.5 days, whereas analytic scoring took 8.5 days. According to Klein et al. (1998), grading a Grade 5 student using an analytic rubric took 17.5 minutes and cost \$7.29, whereas a holistic rubric took 6.4 minutes and cost \$2.67. Analytic rubrics are more time-consuming than holistic rubrics.

Rubrics should be reliable and valid, reliability ensures score consistency across repeated measurements, while validity assesses how well scores reflect the underlying variable of interest (Stellmack et al., 2009). Reliability is a prerequisite for valid assessments, but reliable assessments are not necessarily valid (Moskal & Leydens, 2000). Restricting assessments for better reliability may reduce validity by not measuring the full scope (Jonsson & Svingby,

2007). Consequently, a student's full potential may not be captured, and the unanticipated potential is likely disregarded (Bennett, 2016). Validity is more critical than reliability, and achieving it without sacrificing reliability is possible (Jonsson & Svingby, 2007).

d) Validity

Validity concerns the alignment of assessment criteria with what is desirable and expected of students (Kavanagh & Luxton-Reilly, 2016). It encompasses content, construct, and criterion validity (Jonsson & Svingby, 2007; Moskal & Leydens, 2000; Reddy & Andrade, 2010).

Content validity ensures rubric criteria align with the assessed content (Reddy & Andrade, 2010). It concerns how well a rubric gauges the content domain and students' responses reflect their knowledge (Moskal & Leydens, 2000). For instance, a mathematics test comprising primarily addition problems provides inadequate evidence of a student's ability to do subtraction, multiplication, and division. Similarly, a history exam with complex sentence structures may unintentionally measure reading comprehension instead of historical knowledge (Moskal & Leydens, 2000). Sometimes, scoring high on isolated factors does not necessarily lead to good writing, or vice versa (Rezaei & Lovorn, 2010).

Construct validity addresses content relevance, representativeness, and criterion-relatedness (Messick, 1995). It reflects whether a rubric describes a single latent factor (Gray et al., 2019) and examines an individual's internal processes. For example, a correct answer resulting from incorrect reasoning lacks construct validity (Moskal & Leydens, 2000).

Criterion validity assesses whether validity evidence from one situation can be generalised to another (Rafilson, 1990). For example, assessments mimicking real working environments examine students' ability to function as professional engineers (Moskal & Leydens, 2000). The assessment's quality relies on identifying the components leading to successful performance in the professional environment (Moskal & Leydens, 2000).

A good assessment should contain all three aspects of validity. A rubric possessing only content validity may properly assess content knowledge but not the thinking process. This may yield unfair and misleading results, demotivating students and leading teachers to incorrect educational decisions. Simply providing a rubric does not guarantee content representativeness, scoring structure fidelity, or generalizability, and it lacks convergent or discriminant evidence to other measures (Jonsson & Svingby, 2007).

A rubric with less than six evaluation criteria is manageable (Liew et al., 2020; Popham, 1997). The benefits thereafter would be minimal. Covill (2012) observed that students considering eleven criteria did not write better (or worse) than those with five criteria. Given the restricted number, rubric criteria should be strategized to cover all key aspects with minimal redundancy.

Three to five performance levels are the most common (Brookhart, 2018). Additional levels are unnecessary. Levels exceeding seven are either not used or collapsed by raters into fewer. Although five-point scales are common, an odd number of categories is not recommended (Peeters et al., 2010). The middle category is often over-selected (Weems & Onwuegbuzie, 2001), which can be problematic for interpretation, a middle-category response could be due to a true perceived average, or rater indecisiveness over the scale. Thus, Peeters et al. (2010) suggested a four-rating scale for measurement.

A rubric should align with the assessment goals, adhering to the curriculum and desired learning outcomes. Other factors, like appropriate indicators of the underlying construct, may be included (Moskal & Leydens, 2000). While assessing answers and reasoning, both should be reflected in the rubric. Constructs that provide evidence of students' underlying processes, such as problem-solving, creativity, writing process, self-esteem, and attitudes, may be incorporated (Moskal & Leydens, 2000). The goal is to estimate an assessment as accurately as possible, if not exactly, at least a better approximation (Kane, 2001).

Group-based assessments may have validity issues since work tends to be subdivided among students (Kavanagh & Luxton-Reilly, 2016), and individual students may put up differing levels of knowledge and commitment (Mark et al., 2006). Students might not meet all the assessment criteria individually (Kavanagh & Luxton-Reilly, 2016), but the entire group usually receives the same ratings. This might over-reward low commitment and under-reward high commitment (Mark et al., 2006). The question arises whether group scores should be individualised and, if so, how to individualise group scores.

A group assessment may comprise all content, construct, and criterion validity. This might not be true if the same group of students are assessed individually, particularly in terms of content validity. Despite technical aspects (content), group work develops skills like decision-making, negotiation, communication, empathy, and delegation (conduct) (Mark et al., 2006). These transferable skills are valuable outside the immediate assessment task or content and support life-long learning, personal development planning, and employability (criterion). Rather than

obsessing over certain content or tasks, the underlying construct is more important in group work. Group work can be effective and positively impact students' 'graduateness' (i.e., criterion validity) (Mark et al., 2006). An assessment capable of facilitating the learning process is still valid.

Another validity issue is indeterminacy. It concerns the inadequacy of a rubric in representing the full complexity of multi-criterion qualitative judgement, leading to distorted grading decisions (Sadler, 2009). It represents the condition where a rubric's proposed solution system cannot completely solve a given problem within its own parameters (Sadler, 2009). For example, when breaking down holistic judgements, no matter how many and how carefully criteria are selected, the full complexity of the qualitative judgement cannot be sufficiently represented (Panadero & Jonsson, 2020). Even if it could, the output would be suboptimal (Sadler, 2009).

3.2. Discussion

Rubrics are assessment tools used to clarify learning goals, design instruction addressing the goals, communicate the goals to students, guide feedback on students' progress towards the goals, and judge final products in terms of the degree to which the goals were met (Andrade, 2005). Like other tools, rubrics have potential and limitations. Their potential should not be overstated, and their limitations should not be ignored. Rubrics can, and in many cases do, fail to ensure sound assessments (Sadler, 2009).

Most institutions mandate rubrics for subjective assessments. Rubrics hold academic programmes more accountable for the quality of academic achievement and graduate learning (Sadler, 2014). Instead of opposing, teachers should focus on increasing competencies and finding ways to overcome the limitations of rubrics. There are negative comments about rubrics. Some critics rely on anecdotal evidence or personal experiences, which have limited value as scientific evidence (Panadero & Jonsson, 2020). Nevertheless, critiques provide insights into various potential shortcomings associated with the use of rubrics.

Rubrics rely on fairness, reliability, and validity for effective grading, but achieving perfection is challenging. Even experienced teachers find creating flawless rubrics difficult. While rubrics are expected to enhance learning, their primary function is to facilitate sound assessment. It is acceptable if rubrics are not explicitly designed for instructional purposes, as long as they do not hinder learning. Rubrics can be constructed differently depending on the assessment's goals. In formative assessment, rubrics are utilised for individualised feedback, self-monitoring, and fostering learning development to enhance a student's competency. Summative assessment, on the other hand, focuses on summarising a student's overall achievement, emphasising fairness, reliability, and validity. Rubrics, whether summative or formative, necessitate examination of reliability and validity indicators (Panadero et al., 2013).

Education is not confined to a specific course, individual, or tool. Also, it takes time to manifest. Rubrics cannot be expected to resolve all educational challenges; effective teacher facilitation remains equally crucial. A well-designed rubric cannot be correct for a poorly designed assessment (Moskal & Leydens, 2000). However, this does not exempt rubrics from the need for proper design and implementation, as many limitations are due to poor construction (Oakleaf, 2008). A rubric, whether good, bad, or ugly, depends on how it is created and used (Andrade, 2005). Ill-designed and/or poorly implemented rubrics can be more harmful than beneficial (Panadero & Jonsson, 2020). Rubrics in the hands of unthinking and insensitive teachers have predictably unimpressive and harmful results (Wilson, 2007).

Reality often diverges from expectations. Teachers are expected to create effective rubrics but many fall short. Even with explicit evaluation criteria, the rating process can remain vague, with teachers relying on overall impressions. Similarly, students are supposed to interpret rubrics accurately, but many struggle to meet expectations. Students are supposed to benefit from using rubrics in peer and self-assessment; since neither counts towards final grades (Andrade, 2005), they might not bother doing it properly. Teachers should recognise that students do not always value the use of rubrics, rubrics' positive effects are not always present, and simply handing out rubrics does not guarantee success (Panadero & Romero, 2014).

A rubric must be aligned with reasonable and respectable standards (fairness); be consistent with the curriculum (validity); and result in similar ratings when used by different people (reliability). Effective rubrics allow teachers to provide individualised and constructive critique in a manageable time frame, but they do not relieve teachers from spending time giving students feedback (Andrade, 2005). Feedback is important as rubric criteria are inherently fuzzy and open to interpretation, and the requirements are invariably expressed in relative rather than absolute terms (Sadler, 2014).

The consensus on whether a rubric should be brief and subjective or detailed and explicit may never be reached. There might not be any universal solution fitting all circumstances. Teachers should choose an approach that best

aligns with their strengths and weaknesses to achieve learning goals. Rubrics need not be overly rigid; flexibility allows teachers to creatively customize their facilitation approaches.

A rubric may not function as expected initially, and this is acceptable, considering that developing rubrics is an ongoing process. Continuously refining rubrics over time can lead to an enhancement in assessment quality. Teachers can consistently monitor, review, and make necessary amendments to the rubrics to ensure their effectiveness. Also, students should be educated to use rubrics properly. They should understand the standard, compare their performance with the standard, and take appropriate actions to close the gap (Sadler, 1989). This is, however, insufficient, as it leads to instrumental learning and "criteria compliance" among students (Panadero & Jonsson, 2020).

The importance of grades and their potential impact on students' career prospects should not be overemphasised (Adedoyin, 2013). Teachers should focus on learning beyond the immediate outcomes (Boud & Soler, 2016). This involves fostering positive learning attitudes; acquiring and integrating new knowledge; extending and refining knowledge; using knowledge to perform meaningful tasks; and developing powerful habits of mind for behavioural regulation and critical and creative thinking (Reddy & Andrade, 2010). Despite learning basic knowledge, students should use the knowledge in authentic, real-life situations and apply advanced cognitive skills such as problem-solving, critical thinking, reasoning, decision-making, communicating, and taking responsibility (Kutlu et al., 2010).

Teachers should be cautious about rubrics. Their overuse may limit students' imagination by focusing on students following, not exploring; emphasising students doing, not understanding; and constraining students within arbitrary boundaries (Chapman & Inman, 2009). This may be an issue if the goal is to equip students with problem-solving and critical-thinking skills. Creating rubrics can be time-consuming, especially for those lacking familiarity or expertise (Oakleaf, 2008). Teachers overcommitting time to rubrics may neglect other crucial responsibilities like lesson planning. With limited time and energy, teachers face challenges in ensuring both productive learning and high-quality evaluation. Striking a balance between the two is essential. Demanding "perfect" rubrics from teachers might not necessarily yield desirable educational outcomes, and excessive demands can dampen teachers' enthusiasm.

4. IMPLICATIONS AND CONTRIBUTIONS

4.1 Research Implication

This study provides valuable insights for educators on the effective use of rubrics in educational assessments. By exploring best practices in rubric design and implementation, the findings can help teachers enhance transparency and fairness in assessments while minimizing potential biases in evaluations. Additionally, the study underscores the importance of flexibility for educators in deciding whether to use rubrics and how to design them in alignment with learning objectives to ensure reliability and validity. The results of this study can also inform educational policy development, offering a foundation for more effective rubric-based assessment practices across various educational levels.

5.2 Research Contribution

This study makes a significant contribution by identifying best practices for using rubrics as assessment tools in education. Through a comprehensive review of existing literature, it provides valuable insights into how rubrics can be optimized to enhance assessment quality, ensure transparency, and improve learning outcomes. The study also offers guidelines for educators on designing rubrics that are reliable, valid, and aligned with learning objectives. These contributions aim to support educators, researchers, and policymakers in implementing rubrics more effectively across diverse educational settings, ultimately enhancing the quality of education and assessment practices.

5. LIMITATIONS AND RECOMMENDATIONS

5.1 Limitations of the Study

This study is limited by its reliance on existing literature, which may not fully capture the most recent developments or practical challenges faced by educators in real-world settings. Additionally, the review is confined to studies and articles published in English, potentially overlooking relevant research in other languages. The findings are based on secondary data, which may not account for the diversity of educational contexts or specific cultural and institutional factors that could influence the effectiveness of rubric implementation. As such, the results

may not be fully generalizable to all educational settings, and further empirical research is needed to validate these findings in different contexts.

5.2 Recommendations for Future Research

Future research should focus on empirical studies exploring the real-world application and effectiveness of rubrics in diverse educational settings. Investigating how rubrics impact student learning outcomes, teacher perceptions, and assessment practices across various cultural and institutional contexts would provide a more comprehensive understanding of their utility. Additionally, research could examine the challenges educators face when implementing rubrics and the strategies they use to overcome these obstacles. Longitudinal studies would also be valuable in assessing the long-term impact of rubric use on educational practices and student achievement. Furthermore, further exploration of the integration of digital tools in rubric design and assessment may provide insights into how technology can enhance rubric-based evaluations.

6. CONCLUSION

Debates about rubrics persist. Advocates for reform recognise both the potential and flaws in rubrics, stressing the need for improvements. Those calling for discontinuation highlight various ways rubrics can fail, arguing that improper use may be worse than not using them at all. Despite the disagreement, both sides agree that rubrics are imperfect. Consensus on good rubrics may never be achieved. The levels of explicitness, subjectivity, and reductivity would vary depending on circumstances, subject to the professional discretion of teachers.

Debates often revolve around optimal evaluation methods rather than fundamental purposes of education. There can be various methods but only one educational goal. A rubric's main function is to support an assessment that drives learning. Ultimately, everything converges on learning. As long as learning occurs, the presence or absence of a rubric is less crucial, not to mention its specific characteristics.

It is time to rethink education. The traditional, restrictive approach may unintentionally encourage rule-following in students, kill creativity, and hinder them from unleashing their full potential. This calls for more daring ideas, envisioning a revamped education system that is more adaptive and inclusive. This includes giving teachers greater autonomy and allowing students to blossom in diverse ways. Ideally, both teachers and students should find joy in the process of teaching and learning.

Teachers could be given the flexibility to decide whether to use rubrics in their classes. If rubrics are chosen, they should be properly designed and implemented. This involves incorporating the good characteristics of rubrics, avoiding common mistakes, navigating within their limitations, and continuously enhancing rubrics to achieve the desired level of fairness, reliability, and validity. While current education strongly encourages the use of rubrics, it should also be acceptable if teachers choose to exclude them, as long as the alternative approaches align with educational goals. In such cases, teachers must consistently monitor, review, and refine these approaches to better achieve the educational goal.

Acknowledgments

The researcher would like to thank colleagues who have provided a lot of input suggestions so that the researcher can complete the research process on time.

Author Contribution Statement

The author declares that this article's entire research and writing process was carried out independently. The author is fully responsible for all data related to this research. No other party has participated as an author or made a significant contribution to the content of this work.

Conflict of Interest Statement

The author declares no conflict of interest regarding this research, which was undertaken to enhance the author's competence.

Ethical Approval Statement

The author declares that this study is a literature review and does not involve human participants, personal data, or any other subjects. Therefore, this study does not require ethical approval from a research ethics committee.

The entire research process was conducted in accordance with academic ethical standards, upholding scientific honesty, integrity, and the ethical use of legitimate sources.

REFERENCES

- Abram, P., Scarloss, B., Holthuis, N., Cohen, E., Lotan, R., & Schultz, S. E. (2002). The use of evaluation criteria to improve academic discussion in cooperative groups. *Asia Pacific Journal of Education*, 22(1), 16-27. <https://doi.org/10.1080/0218879020220103>
- Adedoyin, C. A. (2013). Debate-proof grades: Experiences & challenges of using a grading rubric in a social welfare policy course. *Journal of Teaching in Social Work*, 33(2), 196-208. <https://doi.org/10.1080/08841233.2013.774304>
- Andrade, H. G. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education*, 4(4), 1-22. https://scholarsarchive.library.albany.edu/etap_fac_scholar/6
- Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27-30. <http://www.jstor.org/stable/27559213>
- Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27(2), 3-13. <https://doi.org/https://doi.org/10.1111/j.1745-3992.2008.00118.x>
- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation* 10(3), 1-11. https://scholarsarchive.library.albany.edu/edpsych_fac_scholar/2
- Baker, E. L., Abedi, J., Linn, R. L., & Niemi, D. (1996). Dimensionality and generalizability of domain-independent performance assessments. *The Journal of Educational Research*, 89(4), 197-205. <http://www.jstor.org/stable/27542035>
- Bennett, C. (2016). Assessment rubrics: Thinking inside the boxes. *Learning and Teaching*, 9(1), 50-72. <https://doi.org/10.3167/latiss.2016.090104>
- Bloxham, S., den-Outer, B., Hudson, J., & Price, M. (2016). Let's stop the pretence of consistent marking: Exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, 41(3), 466-481. <https://doi.org/10.1080/02602938.2015.1024607>
- Bolton, F. C. (2006). Rubrics and adult learners: Andragogy and assessment. *Assessment Update*, 18(3), 5-6. <https://api.semanticscholar.org/CorpusID:113921569>
- Boud, D. (1989). The role of self-assessment in student grading. *Assessment & Evaluation in Higher Education*, 14(1), 20-30. <https://doi.org/10.1080/0260293890140103>
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-413. <https://doi.org/10.1080/02602938.2015.1018133>
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3(22), 1-12. <https://doi.org/10.3389/feduc.2018.00022>
- Bryant, C. L., Maarouf, S., Burcham, J., & Greer, D. (2016). The examination of a teacher candidate assessment rubric: A confirmatory factor analysis. *Teaching and Teacher Education*, 57, 79-96. <https://doi.org/https://doi.org/10.1016/j.tate.2016.03.012>
- Chan, Z., & Ho, S. (2019). Good and bad practices in rubrics: The perspectives of students and educators. *Assessment & Evaluation in Higher Education*, 44(4), 533-545. <https://doi.org/10.1080/02602938.2018.1522528>
- Chapman, V. G., & Inman, M. D. (2009). A conundrum: Rubrics or creativity/metacognitive development? *Educational Horizons*, 87(3), 198-202. <https://files.eric.ed.gov/fulltext/EJ849020.pdf>
- Cohen, E. G., Lotan, R. A., Abram, P. L., Scarloss, B. A., & Schultz, S. E. (2002). Can Groups Learn? *Teachers College Record*, 104(6), 1045-1068. <https://doi.org/10.1111/1467-9620.00196>
- Covill, A. E. (2012). College students' use of a writing rubric: Effect on quality of writing, self-efficacy, and writing practices." *Journal of Writing Assessment*, 5(1), 1-16. <https://escholarship.org/uc/item/5f79g6kz>
- Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, 31(2), 115-128. <https://doi.org/10.1007/s10755-006-9012-x>
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self-assessments. *Assessment & Evaluation in Higher Education*, 11(2), 146-166. <https://doi.org/10.1080/0260293860110206>
- Galván-Sánchez, I., Verano-Tacoronte, D., González-Betancor, S. M., Fernández-Monroy, M., & Bolívar-Cruz, A. (2017). Assessing oral presentation skills in electrical engineering: Developing a valid and reliable rubric. *International Journal of Electrical Engineering & Education*, 54(1), 17-34. <https://doi.org/10.1177/0020720916659501>

- Gray, J. S., Connolly, J. P., & Brown, M. A. (2019). Measuring intercultural knowledge and competence in college essays: Does a performance-based rubric have construct validity? *Studies in Educational Evaluation*, 62, 142-148. <https://doi.org/https://doi.org/10.1016/j.stueduc.2019.05.007>
- Green, R., & Bowser, M. (2006). Observations from the Field. *Journal of Library Administration*, 45(1-2), 185-202. https://doi.org/10.1300/J111v45n01_10
- Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509-1528. <https://doi.org/10.1080/0950069022000038268>
- Hendry, G. D., Armstrong, S., & Bromberger, N. (2012). Implementing standards-based assessment effectively: incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education*, 37(2), 149-161. <https://doi.org/10.1080/02602938.2010.515014>
- Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27(1), 53-64. <https://doi.org/10.1080/03075070120099368>
- Huang, S. C. (2012). Like a bell responding to a striker: Instruction contingent on assessment. *English Teaching: Practice and Critique*, 11(4), 99-119. <https://api.semanticscholar.org/CorpusID:56007690>
- Jeong, H. (2015). Rubrics in the classroom: Do teachers really follow them? *Language Testing in Asia*, 5(1), 6. <https://doi.org/10.1186/s40468-015-0013-5>
- Johnson, C. S., & Gelfand, S. (2013). Self-assessment and writing quality. *Academic Research International*, 4(4), 571-580. [http://www.savap.org.pk/journals/ARInt./Vol.4\(4\)/2013\(4.4-60\).pdf](http://www.savap.org.pk/journals/ARInt./Vol.4(4)/2013(4.4-60).pdf)
- 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/https://doi.org/10.1016/j.edurev.2007.05.002>
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342. <https://doi.org/10.1111/j.1745-3984.2001.tb01130.x>
- Kavanagh, S., & Luxton-Reilly, A. (2016). Rubrics used in peer assessment. Proceedings of the Australasian Computer Science Week Multiconference, Canberra, Australia. <https://doi.org/10.1145/2843043.2843347>
- Klein, S. P., Stecher, B. M., Shavelson, R. J., McCaffrey, D., Ormseth, T., Bell, R. M., Comfort, K., & Othman, A. R. (1998). Analytic versus holistic scoring of science performance tasks. *Applied Measurement in Education*, 11(2), 121-137. https://doi.org/10.1207/s15324818ame1102_1
- Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12(1), 26-43. <https://doi.org/https://doi.org/10.1016/j.asw.2007.04.001>
- Kutlu, Ö., Bilican, S., & Yıldırım, Ö. (2010). A study on the primary school teachers' attitudes towards rubrics with reference to different variables. *Procedia - Social and Behavioral Sciences*, 2(2), 5398-5402. <https://doi.org/https://doi.org/10.1016/j.sbspro.2010.03.880>
- Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), 157-172. <https://doi.org/10.1080/03075079812331380364>
- Li, J., & Lindsey, P. (2015). Understanding variations between student and teacher application of rubrics. *Assessing Writing*, 26, 67-79. <https://doi.org/10.1016/j.asw.2015.07.003>
- Liew, C. P., Puteh, M., & Hamzah, S. H. (2020). Comparative study of engineering design project assessment rubrics to address the Washington Accord's complexity attributes. *ASEAN Journal of Engineering Education*, 4(1), 71-94. <https://doi.org/10.11113/ajee2020.4n1.21>
- Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: implications for training. *Language Testing*, 12(1): 54-71. <https://doi.org/10.1177/026553229501200104>
- Malouff, J. (2008). Bias in grading. *College Teaching*, 56(3), 191-192. <https://doi.org/10.3200/CTCH.56.3.191-192>
- Mark, R., George, H., & Alan, C. (2006). Individualising students' scores using blind and holistic peer assessment. *Engineering Education*, 1(1), 50-60. <https://doi.org/10.11120/ened.2006.01010050>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment, Research, and Evaluation*, 7(3), 1-5. <https://doi.org/10.7275/a5vq-7q66>
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research, and Evaluation*, 7(7), 1-6. <https://doi.org/10.7275/q7rm-gg74>
- Murillo, F. J., & Hidalgo, N. (2017). Students' conceptions about a fair assessment of their learning. *Studies in Educational Evaluation*, 53: 10-16. <https://doi.org/10.1016/j.stueduc.2017.01.001>

- Oakleaf, M. (2008). Dangers and opportunities: A conceptual map of information literacy assessment approaches. *portal: Libraries and the Academy*, 8(3): 233-253. <https://doi.org/10.1353/pla.0.0011>
- Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education*, 21(3), 239-250. <https://doi.org/10.1080/0260293960210304>
- Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, 30: 100329. <https://doi.org/https://doi.org/10.1016/j.edurev.2020.100329>
- Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, 21(2), 133-148. <https://doi.org/10.1080/0969594X.2013.877872>
- Panadero, E., Romero, M., & Strijbos, J. W. (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. *Studies in Educational Evaluation*, 39(4), 195-203. <https://doi.org/https://doi.org/10.1016/j.stueduc.2013.10.005>
- Parke, C. S. (2001). An approach that examines sources of misfit to improve performance assessment items and rubrics. *Educational Assessment*, 7(3), 201-225. https://doi.org/10.1207/S15326977EA0703_02
- Peeters, M. J., Sahloff, E. G., & Stone, G. E. (2010). A standardized rubric to evaluate student presentations. *Am J Pharm Educ* 74(9), 171. <https://doi.org/10.5688/aj7409171>
- Pepper, M. B., & Pathak, S. (2008). Classroom contribution: What do students perceive as fair assessment? *Journal of Education for Business*, 83(6), 360-368. <https://doi.org/10.3200/JOEB.83.6.360-368>
- Petkov, D., & Petkova, O. (2006). Development of scoring rubrics for IS Projects as an assessment tool. Issues in *Informing Science and Information Technology*, 3, 499-510. <https://doi.org/10.28945/910>
- Popham, W. J. (1997). What's wrong - and what's right - with rubrics. *Educational Leadership*, 55(2), 72. <https://eric.ed.gov/?id=EJ552014>
- Rafilson, F. (1990). The case for validity generalization. *Practical Assessment, Research, and Evaluation*, 2(13), 1-3. <https://doi.org/10.7275/b51w-5m96>
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448. <https://doi.org/10.1080/02602930902862859>
- Reynolds-Keefer, L. (2010). Rubric-referenced assessment in teacher preparation: An opportunity to learn by using. *Practical Assessment, Research, and Evaluation*, 15(8): 1-9. <https://doi.org/10.7275/psk5-mf68>
- Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing*, 15(1), 18-39. <https://doi.org/10.1016/j.asw.2010.01.003>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144. <https://doi.org/10.1007/BF00117714>
- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159-179. <https://doi.org/10.1080/02602930801956059>
- Sadler, D. R. (2014). The futility of attempting to codify academic achievement standards. *Higher Education*, 67(3), 273-288. <https://doi.org/10.1007/s10734-013-9649-1>
- Sambell, K., McDowell, L., & Brown, S. (1997). But is it fair?: An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23(4), 349-371. [https://doi.org/https://doi.org/10.1016/S0191-491X\(97\)86215-3](https://doi.org/https://doi.org/10.1016/S0191-491X(97)86215-3)
- Scott, S., Webber, C. F., Lupart, J. L., Aitken, N., & Scott, D. E. (2014). Fair and equitable assessment practices for all students. *Assessment in Education: Principles, Policy & Practice*, 21(1), 52-70. <https://doi.org/10.1080/0969594X.2013.776943>
- Simon, M., & Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research, and Evaluation*, 7(18), 1-4. <https://doi.org/10.7275/bh4d-me80>
- Stellmack, M. A., Konheim-Kalkstein, Y. L., Manor, J. E., Massey, A. R., and Schmitz, J. A. P. (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teaching of Psychology*, 36(2), 102-107. <https://doi.org/10.1080/00986280902739776>
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation*, 9(4), 1-11. <https://doi.org/10.7275/96jp-xz07>
- Sundeen, T. H. (2014). Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21, 74-88. <https://doi.org/https://doi.org/10.1016/j.asw.2014.03.003>
- Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research, and Evaluation*, 9(2), 1-7. <https://doi.org/10.7275/jvtv-wg68>

Van Helvoort, A. A. J. (2012). How adult students in information studies use a scoring rubric for the development of their information literacy skills. *The Journal of Academic Librarianship*, 38(3), 165-171. <https://doi.org/https://doi.org/10.1016/j.acalib.2012.03.016>

Weems, G. H., & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. *Measurement and Evaluation in Counseling and Development*, 34(3), 166-176. <https://doi.org/10.1080/07481756.2002.12069033>

Williams, L., & Rink, J. E. (2003). Teacher competency using observational scoring rubrics. *Journal of Teaching in Physical Education*, 22: 552-572. <http://journals.humankinetics.com/jtpe-contents>

Wilson, M. 2007. Why I won't be using rubrics to respond to students' writing. *The English Journal*, 96(4), 62-66. <https://doi.org/10.2307/30047167>

Zhang, B., Xiao, Y., & Luo, J. (2015). Rater reliability and score discrepancy under holistic and analytic scoring of second language writing. *Language Testing in Asia*, 5(1), 5. <https://doi.org/10.1186/s40468-015-0014-4>

Article Information

Copyright holder: © Ling, J. H. (2025) First Publication Right: Indonesian Journal of Innovative Teaching and Learning Article info: DOI: https://doi.org/10.64420/ijitl.v2i1.197 Word Count: 9748	Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of AEDUCIA and/or the editor(s). AEDUCIA and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content. This Article is licensed under the CC-BY-SA 4.0
--	---