

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Exploring Postgraduate Students' Experience with Rubric-referenced Assessment: Limitations and Solutions

Liang Jing Teh¹, Su Luan Wong^{1*}, Mas Nida Md Khambari¹, Rahmita Wirza O. K. Rahmat² and Sai Hong Tang³

¹Department of Science and Technical Education, Faculty of Educational Studies, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

²Department of Multimedia, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

³Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

ABSTRACT

Despite the various benefits of rubric-referenced assessment (RRA), multiple studies have revealed its potential pitfalls. Given the scarcity of research on the limitations of RRA and its solutions in the context of Malaysian postgraduates, this study explores the limitations of RRA and proposes potential strategies for improvement from the perspective of postgraduate students in Malaysia. The study adopted a case study qualitative approach and the Activity Theory as the theoretical framework. Five Malaysian postgraduate students provided their responses via two focus group discussions. The participants highlighted that rubrics may stifle creative self-expression, cause inconsistency in scoring, cause confusion among students, and be limited in catering to various learner needs. To address these issues, the participants recommended that the instructor allocate time for students to understand the rubric and engage in discussion about its content. They also proposed flexibility in rubrics to accommodate revisions based on student feedback and implementing

ARTICLE INFO

Article history: Received: 04 January 2024 Accepted: 20 December 2024 Published: 30 April 2025

DOI: https://doi.org/10.47836/pjssh.33.2.03

E-mail addresses: tehliangjing@gmail.com (Liang Jing Teh) suluan@upm.edu.my (Su Luan Wong) khamasnida@upm.edu.my (Mas Nida Md Khambari) rahmita@upm.edu.my (Mahmita Wirza O. K. Rahmat) saihong@upm.edu.my (Sai Hong Tang) * Corresponding author scoring calibration sessions or training to maintain scoring consistency. Other suggestions comprised prioritising inclusive assessments, tailoring rubrics for different learner profiles, including specific numerical indicators in rubric descriptions, using a holistic rubric, and providing feedback to students according to the rubric.

Keywords: Analytic rubrics, areas of consideration, focus group discussion, higher education, limitations, postgraduate students, rubric-referenced assessment

INTRODUCTION

Rubric-referenced assessment (RRA) has emerged as a widely adopted framework that offers a structured and transparent approach to evaluating university students' work (Brookhart, 2018; Cockett & Jackson, 2018; Noh et al., 2021). Despite the recognition that RRA has garnered, it is essential to critically investigate its limitations and challenges. Previous studies have shown that postgraduate students may perceive rubrics to be limited in clarifying the assessor's expectations (Kite & Phongsavan, 2017) and evaluating complex thinking skills (Bennett, 2016). Inconsistent interpretation among different users of rubrics could also affect the scoring reliability of assessments (Venning & Buisman-Pijlman, 2013). Bennett (2016) also claimed that RRA might reduce learners' creativity and might not aid the feedback loop between students and instructors.

Several scholars remarked that instructors should provide genuine feedback on learners' work rather than relying on rubrics (Bennett, 2016). Assessors were also advised to be formally trained to improve scoring consistency (Postmes et al., 2023), simplify the grading system to include broader learning outcomes, as well as have iterative feedback with the students (Kite & Phongsavan, 2017; Venning & Buisman-Pijlman, 2013). Bennett (2016) also suggested reducing the stakes value of academic performance and encouraging intellectual risk-taking.

Malaysian universities have initiated alternative assessments such as RRA since the early 2000s (Darus et al., 2003). Since then, RRA has become a prevalent method for evaluating tertiary-level students' performance in Malaysia (Bukhari et al., 2021; Saeed et al., 2019). This is in line with the nation's plan to replace conventional assessments with alternative assessments that could develop 21st-century skills among university students (Ministry of Higher Education Malaysia, 2021). It is vital to understand the postgraduates' perspective on the use of RRA to foster a more effective integration of RRA in higher education.

Nevertheless, a knowledge gap exists within the literature regarding RRA's limitations and its potential solutions, especially from the perspectives of Malaysian postgraduate students. This is because the stated limitations and potential solutions of RRA were not in the context of Malaysian postgraduates. Also, most existing literature on RRA only reported findings among undergraduate students, and only a few studies focus on investigating RRA's limitations and solutions (Brookhart, 2018; Postmes et al., 2023). Hence, this study aims to address this gap by gaining deeper insights into Malaysian postgraduates' perspectives on the limitations of RRA and the potential solutions.

LITERATURE REVIEW

Rubric-referenced Assessment

RRA is an evaluation method that uses a predefined set of criteria or standards to assess and grade students' work. It provides a structured framework for evaluating student performance, offering clear guidelines and expectations for what constitutes highquality work. Different practitioners and academicians have had different definitions of rubrics throughout the past, but according to Popham (1997), rubrics have three main elements. The first is specificity, and the rubric should clearly and specifically define the criteria that will be used to evaluate student performance. The second element is inter-rater reliability; the criteria and categories in the rubric must be explicit enough to ensure that different evaluators will similarly apply them. Finally, the third element is flexibility. Rubrics must be flexible enough to accommodate numerous levels of student performance and be available for subjective judgment. Rubrics can be designed as either analytic or holistic rubrics (Brookhart, 2013). An analytic rubric has several assessment criteria in separate rows, which are described in columns according to the rating scale. In a holistic rubric, all criteria are combined into a single overall performance description and arranged along rows of the rating scale (Popham, 1997).

RRA is widely used in HE as a valuable tool for evaluating student performance (Kite & Phongsavan, 2017; Venning & Buisman-Pijlman, 2013). Previous studies have found that the benefits of RRA include clarifying instructor's expectations (Kite & Phongsavan, 2017; Pérez-Guillén et al., 2022; Venning & Buisman-Pijlman, 2013), promoting consistency and transparency in grading (Andrade & Du, 2005; Chowdhury, 2019; Venning & Buisman-Pijlman, 2013), enhancing feedback and assessment literacy (Chowdhury, 2019; Postmes et al., 2023), allowing students to engage in self-reflection and self-assessment (Anandi & Zailaini, 2019; Bukhari et al., 2021; Saeed et al., 2019; Venning & Buisman-Pijlman, 2013), supporting differentiation and individualisation in assessing student work (Chowdhury, 2019; Holmstedt et al., 2018; Kite & Phongsavan, 2017), as well as facilitating students' self-assessment (Anandi & Zailaini, 2019; Fraile et al., 2017; Jonsson, 2014; Oakleaf, 2009; Venning & Buisman-Pijlman, 2013).

Navigating the Potential Drawbacks of Rubric-referenced Assessment

Despite the aforementioned benefits of RRA in higher education, previous work also revealed the potential drawbacks of RRA. Notably, a qualitative study done by Kite and Phongsavan (2017) offers a crucial student-centred insight, suggesting that even though university students were provided with rubrics, they still lack tacit knowledge of the underlying criteria in the rubrics. This discrepancy between the lecturer's expectations and students' understanding of assessment rubrics can result in substantial disparities in grasping the essence of assessment rubrics. Panadero and Romero (2014) did a comparative study that comprised 218 prospective teachers and concluded that the students who utilised rubrics reported more stress and a lower tendency to self-regulate their learning, potentially hindering their autonomy and self-guided academic growth.

These student-centred critiques were supported by the opinions of Bennett (2016)

and Sadler (2014), who questioned the pedagogical implications. They contended the importance of nurturing independent judgment and the discernment of quality work among students rather than merely adhering to pre-set criteria in rubrics. A qualitative study among 12 postgraduate participants and supervisors conducted by Venning and Buisman-Pijlman (2013) argued that different users of rubrics would have different interpretations of rubrics. Bennett (2016) and Sadler (2014) also expressed the same sentiments, indicating that rubrics could be limited in assessing subjective work like essays or artwork, which would conjure inconsistency in scoring. A similar concern was raised by Sitorus (2020), who reasoned that too general rubrics may result in inconsistent learners' results. On the other hand, too detailed rubrics may not accommodate educators' subjectivity and may overlook abstract areas such as originality and creativity shown by the learners.

The purpose of HE should be to develop students of higher-order thinking skills (Bennett, 2016). Providing students with transparent assessment processes and criteria could lead them to focus on shallow strategies in complying with the criteria (Bearman & Ajjawi, 2021; Venning & Buisman-Pijlman, 2013). With regards to the drawbacks of using rubrics, Matshedisho (2020) suggested that lecturers should discuss the rubrics with students because it not only closes the gap between the lecturer's expectations and students' interpretation of rubrics but students' views can also be taken in to improve the implementation of the rubric. Mok and Toh (2015) also stressed the importance of ensuring a shared understanding of the rubrics among students and assessors. This can be achieved through peer and self-assessment practices, where students evaluate themselves using the same rubrics, engage in discussions and negotiate meaning with each other and their assessors. This shared understanding ensures that both students and assessors have a common yardstick of measurement (Mok & Toh, 2015).

Furthermore, Shadle et al. (2012) also recognise the importance of considering learners' different backgrounds and the specific skills and knowledge required in different disciplines. By tailoring rubrics to the specific needs of different learner groups, educators can ensure that the assessment is aligned with the learners' backgrounds and provides a more accurate reflection of their abilities (Shadle et al., 2012). By analysing 313 research skill rubrics used in both formative and summative assessments among postgraduates, Postmes et al. (2023) concluded that training programs for developing specific rubric criteria and providing feedback based on these criteria are crucial. The authors also stated that flexibility in applying rubrics should be provided to instructors so that personalised feedback can be given to the students. Venning and Buisman-Pijlman (2013) also suggested that formal training among instructors could improve scoring consistency.

Sadler (2014) also recommended having a group of competent judges classify a body of student work and elaborate exactly why such a classification was reached. When an agreement is warranted, this would concretise the description of criteria as the standards can be referred to as a concrete object. In other words, a sounddesigned rubric is inadequate, and it must be followed by appropriate application to reap the benefits of assessment rubrics (Panadero & Jonsson, 2020). Additionally, iterative feedback loops between instructors and students allow rubrics to be improved through feedback and enable students to understand the rubric's criteria and their performance (Kite & Phongsavan, 2017; Venning & Buisman-Pijlman, 2013).

The lack of relevant studies in the context of Malaysian postgraduates, together with the diverse viewpoints of RRA among university students, indicates a gap in understanding the limitations of RRA and the potential resolutions perceived by postgraduate students in Malaysia. This warrants a deeper inquiry into the potential risks of using RRA and the ways of mitigating them, especially within the context of Malaysian postgraduate students. Thus, this study aims to explore postgraduate students' perspectives on the limitations of RRA and its areas of consideration for improvement. As such, the following presents the research objectives that emerged from the existing literature and theoretical framework:

- To explore postgraduate students' perspectives on the limitations of RRA.
- 2. To explore postgraduate students' perspectives on the areas of consideration in improving RRA.

THEORETICAL FRAMEWORK Activity Theory

The Activity Theory (AT) by Vygotsky (1978) was adopted as the theoretical foundation for this study to gain a deeper understanding of the postgraduate students' experience with RRA. AT is a socio-cultural framework which emphasises the dynamic association between subjects, mediating tools, and the object, which leads to an outcome (Vygotsky, 1978). The subject in an activity system is the individual or group whose perspective is adopted (Engeström, 1993). The term "object" refers to the subject matter or problem space in which the activity is directed, and it serves as the basis that is shaped or transformed into outcomes through tools (Engeström, 1993). Tools play a mediating role in the activity's object and contribute to the transformation of the object into an outcome. This transformation can be either anticipated or unforeseen, as the mediating tools can either facilitate or impede the progress of the activity (Engeström, 1993; Vygotsky, 1978).

Figure 1 illustrates the theoretical framework of this study. In the context of this study, the activity system was used to implement RRA among the postgraduate students. This activity system was analysed using the AT as a framework. The subject, object, and mediating tools were the postgraduate students, academic assessments, and rubrics accordingly. The postgraduate students (subject) interact with the academic assessments (object) under the guidance of the rubrics (mediating tools). In this process, the students could have various experiences with RRA (outcome), such as their experiences in completing the assessments and the perceived clarity of the instructor's expectations. According to AT (Vygotsky, 1978), the integration of RRA might enhance or restrain the students' experience in completing their academic assessments. For instance, the clearly defined description of rubrics could enhance the student's understanding of the instructor's expectations and improve their performance. On the contrary, their performance may be restrained if the rubrics are ambiguous and rigid.

Given the relatively lower number of previous studies on the potential risks of RRA (Bearman & Ajjawi, 2021; Bennett, 2016; Kite & Phongsavan, 2017; Matshedisho, 2020; Panadero & Romero, 2014; Sadler, 2014; Sitorus, 2020; Venning & Buisman-Pijlman, 2013) as compared to the number of previous studies which reported the benefits of RRA (Andrade & Du, 2005; Chowdhury, 2019; Fraile et al., 2017; Holmstedt et al., 2018; Jonsson, 2014; Kite & Phongsavan, 2017; Oakleaf, 2009; Pérez-Guillén et al., 2022; Saeed et al., 2019; Venning & Buisman-Pijlman, 2013), this study focuses on the potential drawbacks of RRA and the ways of overcoming them.

METHODOLOGY

Participants and Procedures

The study was conducted in a Malaysian public university. The participants of this study were selected through purposive criterion sampling as it allowed the researchers to select participants who had experience relevant to the research objectives (Patton, 2014). The inclusion criteria of the participants of this study were (1) postgraduate students of the institution and (2) postgraduate students who have completed the course that incorporated RRA. On the other hand, the exclusion criteria were (1) any students who were not pursuing a postgraduate programme of study and (2) postgraduate students who did not undergo the course which utilised RRA. The postgraduate course that applied



Figure 1. Theoretical framework based on the Activity Theory

for RRA was Educational Technology. One instructor solely implemented it throughout the semester.

Table 1 illustrates the demographics of the participants who have experienced the course, which integrated RRA, and voluntarily participated in this study. All of them were enrolled in the Master of Education in Instructional Technology and Innovation program. They consist of one female and four male postgraduate students from the institution. All the respondents were from an educational field except one that is from the army training field and another from an audio engineering background. Their names have been modified to keep respondents' personal information confidential.

This study employed two focus group discussions (FGDs) to obtain deep insights into students' views of rubrics. Compared to a one-on-one interview, it creates a safer space for respondents to share their opinions and feelings (Krueger, 2014). A FGD protocol was developed by the first author and was evaluated by the other authors to ensure the credibility of the FGD questions. Semi-structured questions were used to facilitate the FGD to acquire accurate information about the research objectives while maintaining the flexibility to explore emerging ideas (Brinkmann, 2014). Semistructured questions combine the strengths of both structured (standardisation) and unstructured questions (flexibility), which allow researchers to elicit rich qualitative data (Brinkmann, 2014). Below are the semi-structured questions included in the FGD protocol:

- 1. In your opinion, what are the limitations of rubrics? Why?
- 2. Do you think rubrics can sometimes be too rigid or too general? Can you provide examples?
- 3. Do you think rubrics can introduce or reinforce biases in grading? If so, how?
- 4. How do you think your understanding of the rubric criteria affects your learning?
- 5. How well do you think rubrics accommodate diverse learning styles and needs?
- 6. Do you have anything else you would like to share regarding the limitations of rubrics?

Respondent	Gender	Age	Career background
FGD1			
Student #1	Male	36	Teacher
Student #2	Male	42	University Media Producer
FGD2			
Student #3	Male	41	Army Officer
Student #4	Male	40	Teacher
Student #5	Female	34	Teacher

Table 1Demographics of respondents

- 7. In your opinion, what are the ways to overcome these limitations?
- 8. How can rubrics be designed to be more flexible or adaptable to different contexts?
- 9. How can we ensure that educators use rubrics effectively?
- 10. How can we ensure that the students fully understand the rubrics?
- 11. How can rubrics be adapted to better suit diverse student populations and learning environments?
- 12. Do you have anything else you would like to share regarding how to overcome the limitations of rubrics?

Additionally, open-ended questions were also prompted spontaneously throughout the FGD so that viewpoints from the respondents are not limited to the researcher's presumption (Creswell & Creswell, 2018). Each FGD took about 35 minutes. At the start of the FGD, the author obtained consent for audio recording from all participants, and a printed copy of the analytical rubric used for the last assignment of the course was given to them to refresh their memory on the rubric used.

Data Analysis

This study employed thematic analysis to analyse data as Terry et al. (2017) stated that this method provides researchers with "accessible and systematic procedures for generating codes and themes from qualitative data" (p. 2). As the respondents conversed in English and Malay language throughout the FGD, the conversation was first transcribed in both languages. The Malay transcript was then translated into English and back-translated to Malay. It was then reviewed by comparing it to the original Malay transcript to ensure language translation accuracy. Three experienced language experts did all the translations and reviews.

Based on the guidelines provided by Creswell and Creswell (2018) as well as Terry et al. (2017), the researchers conducted a thematic analysis of the transcript by first familiarising themselves with the data. The researchers obtained an overall impression of the discussion by reading and rereading the transcripts while keeping the research objectives in mind. Meaning units were then formed by dividing the transcript into smaller parts. Subsequently, the meaning units were condensed into shorter sentences while retaining the core meaning. This was followed by coding, whereby each condensed meaning unit was assigned a code. After coding, all related or repeated codes were grouped into categories about the research questions. The categories and codes were interpreted inductively.

Francis et al. (2010) suggested four principles for reaching data saturation: (1) stating the initial analysis sample, which is the minimum sample size for the first round of data analysis; (2) specifying the stopping criterion, which refers to the sample size to be subsequently included until ideas or themes stopped emerging; (3) the data analysis should be conducted by at least two independent researchers; and (4) to report the data saturation methods and findings.

The initial analysis sample of this study was set to at least two participants for the first FGD, as Bukhari et al. (2021) qualitative study regarding RRA also involved two participants. Nonetheless, the researchers included three more participants for the second FGD as the stopping criterion. No new themes emerged from the second FGD. Therefore, the researchers stopped collecting and analysing data after the second FGD. Although data analysis was only conducted by the first author, the process of the thematic analysis (from meaning units to categories) was tabulated by the first author and validated by the remaining authors. The researcher also conducted member-checking, whereby the participants validated the interpretations by the researcher to ensure there were no discrepancies between the intended meaning and the interpretations.

The inter-rater reliability (R) of this study was assessed according to the formula suggested by Marques and McCall (2005):

R = (total number of agreements) / (total number of observations) \times 100

To identify findings with similar depth, both authors (inter-raters) set the number of listed categories to 13 and allocated seven days to analyse the data (Marques & McCall, 2005). Among the 13 categories derived by both authors, 10 similar categories emerged. Based on Hamilton et al. (2000), the confusion matrix of this study is presented in Table 2.

Thus, the R of this study is 76.92%. The R was calculated as follows:

$$R = 10 / 13 \times 100 = 76.93\%$$

According to Marques and McCall (2005), the accuracy rate (AC) of this study is 72.72%, which was calculated as follows:

$$AC = (10+6) / (10+3+3+6) \times 100$$

= 72.72%

As the R (76.93%) is higher than the AC (72.72%) of this study, the interrater reliability of this study is considered established. The authors then organised the common categories into four themes, with two themes for each research objective.

FINDINGS AND DISCUSSION

Student's Perspective on the Limitations of Rubrics

One of the objectives of this study is to explore postgraduate student's angle on the limitations of RRA. Overall, the respondents pointed out four constraints of the rubric from their experience, which

Table 2Confusion matrix of this study

		Inter-rater 2	
		Agree	Disagree
Inter-rater 1	Agree	10	3
	Disagree	3	6

are (1) limitation in self-expression, (2) inconsistency in scoring, (3) confusion due to the nuances in descriptions, and (4) lack of flexibility in adapting to learners' needs.

Impact of Standardisation on Autonomy Among Students and Educators

During the FGD, the respondents stated that using rubrics in assessing students' performance can limit students from expressing themselves freely and creatively.

"With the rubrics, I couldn't express my own opinion. I am like limited by the rubrics, if I do something different from the rubrics, I would be worried that my marks will be affected. So, I just follow" (Student #3)

This finding highlights the tension between structured assessment and the need for creative autonomy. While rubrics are intended to standardise evaluation, they can inadvertently limit students' ability to explore and express original ideas. This is especially so for students pressured to excel in the assessments. They felt "worried that" their "marks would be affected" if they did not comply with the pre-set criteria despite not fully resonating with them. Although this differs from the findings by Bukhari et al. (2021) and Saeed et al. (2019), it is in line with the critique by Bennett (2016) and Bearman and Ajjawi (2021). Sadler (2014) stated that tertiary students should cultivate autonomous thinking rather than just complying. The phrase "I just follow" that occurred during the FGD suggests that learners may be prone to diminished selfleadership and autonomy when they comply with the pre-determined criteria (Venning & Buisman-Pijlman, 2013).

Apart from the above, one respondent drew from his teaching experience and spoke about the rigidity of rubrics, especially when they are executed on a larger scale.

"When it is implemented on a bigger scale, we [educators] are forced to follow it although we do not agree with it. For example, in my English writing class, is the language or the story more important? I used to prioritise my students' stories. This is what I thought was right but the main criteria to score based on the rubrics is the language." (Student #1)

The respondent further clarified his view by relating to the rubric during the FGD.

"Like for this [rubric for] infographics, is the design more important? Or is it the content? Different lecturers may have different views. So, it is like beyond our [educators'] capabilities to change it. There is a possibility that rubrics would clash with what the instructor feels most suitable for the students." (Student #1)

This rigidity suggests that the pre-set criteria in a centralised rubric could cause a dilemma among the instructors. They are forced to comply with the rubric, and it is beyond their authority to choose to focus on the learning needs of students. In other words, students may focus on meeting predefined criteria when the assessment tools provide structure, which could stifle students' creativity (Bearman & Ajjawi, 2021; Sadler, 2014). However, students may struggle to understand their instructor's expectations and experience anxiety if the success criteria of an assessment are not provided (Pérez-Guillén et al., 2022). This is congruent with the research done by Sitorus (2020), which revealed the challenges in striking the right balance between a rubric that is too general (leading to inconsistent learner results) and one that is overly detailed (neglecting educators' subjectivity and learners' originality and creativity).

Viewing this finding through the lens of AT, the engagement between the postgraduate students (subject) and the academic assessments (object) was affected by the rubrics (mediating tools). While the rubrics aimed to clarify the success criteria of the assignments, they inadvertently constrained students' ability to express original ideas and conflicted with instructors' professional judgment. This could lead to frustration among the students' overall experience with RRA (outcome). This resonates with Vygotsky's (1978) AT that while the tools may be helpful, they can also impose limitations on the subject. This indicates a contradiction within the activity system that needs to be addressed to support students' creativity and learning.

Challenges in Fair and Consistent Assessment

Even with the use of rubrics, inconsistency in final marks persists among learners; a

few of the respondents shared this during the FGD.

"For me, the rubric also has a weakness in terms of scoring. For example, subjects like research methods have many classes, but the learning content and assessments are the same. The lecturer teaching class A and another lecturer teaching class B score their students differently although they are using the same rubric. And from there, we can see the scoring is different." (Student #2)

"Maybe the lecturers' interpretation of the rubric is different." (Student #3)

The inconsistency in scoring points to the potential for subjective interpretation of rubrics among different lecturers, which can undermine the fairness and reliability of assessments. This result is in contrast with the inferences made by Andrade and Du (2005), but it is in parallel with the study conducted by Venning and Buisman-Pijlman (2013), Bennett (2016), and Sadler (2014), which concluded that different interpretations of rubrics could result in inconsistent scoring of assessments.

However, it is pertinent to keep in mind that similarity does not imply consistency, as separate groups of students may have different levels of performance. During the FGD, remarks such as "*The lecturer teaching class A and another lecturer teaching class B score their students differently although they are using the same rubric*" imply that the participants may not be aware of the difference between consistency and similarity. Also, it is important not to misunderstand the intent of rubrics. The aim of rubrics is not to make the score the same but to make scoring more transparent and consistent (Kite & Phongsavan, 2017; Venning & Buisman-Pijlman, 2013).

Moreover, one respondent raised concerns about the challenges she faced in understanding the subtle difference between the descriptions across the rating scale, and another respondent chipped in actively.

"It is hard to justify the difference between the marks because the mark is between 1 and 20, right, so I cannot justify the difference between 16 and 20. It is a very small difference, but the result will be different. Sometimes I think, I can get 16–20, but somehow, in Prof.'s view, I did not hit the category. So sometimes I don't see the difference." (Student #5)

"Sometimes, it is more difficult to answer the questions of rubrics than the actual assignment question. Like Student #5 said, the rubric is too general." (Student #3)

This suggests that even at a postgraduate level, students would still be confused by the specifiers in the rubrics. This observation is in contradiction with the findings by Pérez-Guillén et al. (2022) and Venning and Buisman-Pijlman (2013). However, it aligns with the study done by Kite and Phongsavan (2017), Matshedisho (2020), and Sitorus (2020), which showed that a gap between students' knowledge of rubrics and lecturers' expectations could persist despite the inclusion of descriptions of criteria. This could elicit additional psychological distress in completing their assessment (Panadero & Romero, 2014).

Another respondent agreed with the above comments but said the confusion would subside with time and experience using rubrics.

"... but this is only during the earlier stage of using rubrics. After using it for some time, I think this problem does not exist." (Student #4)

This signifies that as students become more experienced in using rubrics, this problem could fade away. This comment is consistent with the findings of the study conducted by Pérez-Guillén et al. (2022). The findings highlight the obstacles to RRA implementation in terms of ensuring fairness across various evaluators and increased cognitive load among students due to complex rubrics' descriptions (Matshedisho, 2020; Sitorus, 2020). This suggests that the rubric descriptors must be carefully designed and articulated so that the disparities between the users' interpretation and the intended meaning of the rubrics' criteria can be minimised. This also underscores the pertinence of continuous effort from both instructors and students throughout the implementation of RRA. The interpretation of rubrics needs to be standardised among instructors to ensure assessments' validity and reliability. In contrast, students need to be given

guidance and time to understand the intended clarity of the rubrics.

From an AT perspective, the subjects in this case are twofold: the instructors and the postgraduate students. The potential inconsistency in students' academic scores and the students' confusion regarding the subtle differences in the rubrics' descriptions signify a subject-tool contradiction. Although the rubrics were intended to enhance consistency in scoring, the instructors' subjective interpretation could result in inconsistent grading. Similarly, while rubrics were designed to clarify success criteria, the nuances in the description may cause discrepancies in the understanding of rubrics among students. In other words, the rubrics (mediating tool) did not function uniformly in the activity system. Consequently, these could lead to an unpleasant learning experience (outcome). However, as students become more familiar with the rubrics over time, the rubrics could become a more effective mediating tool. This could help reduce the aforementioned contradictions.

Student's Perspective on the Areas of Consideration

When prompted on how to overcome the aforementioned limitations, all the participants talked at length about their suggestions for improving RRA. The suggested solutions include implementing RRA with collaboration and feedback among instructors and students and customising rubric design for inclusive assessments.

Collaborative and Feedback-driven Rubric Implementation

A few respondents recommended that the educators should be trained in applying the rubrics and that the voices of students should be heard during RRA implementation.

"For me, this is for the lecturers, especially for a subject that has multiple lecturers, they should examine their understanding of the rubrics, see if it is the same or not." (Student #2)

"Maybe we can give the lecturers a centralised training... For me it is balancing the weightage for the categories. One way is you can discuss with the students." (Student #1)

This indicates that lecturers should have a scoring calibration session or centralised training among themselves if the same rubric is to be utilised by multiple lecturers. This can increase the consistency in scoring between different lecturers. This comment resonates with the suggestion provided by Postmes et al. (2023), Venning and Buisman-Pijlman (2013), and Sadler (2014), stating that the judges should exchange their evaluations to concretise the rubric's criteria based on concrete students' work. As the assessment rubric is subjective, its design and implementation should be flexible enough to adapt to the feedback received or via reflection postassessment.

The respondents also shared enthusiastically about the importance of giving time to students to digest the information in the rubric and having a discussion together with the lecturer.

"If I can suggest, it would be even better if we [students] were allowed to go back and read through everything, and next week, we would discuss the rubrics and questions. This gives us time to digest. This is because even if you give me one hour, I cannot fully understand [the rubric and the question], and when the lecturer asks us on the spot, "Do you understand?" we will just be like, "Ya, ya" (agreeing]." (Student #3)

"Or explain why you are giving me this. So, as in, like you want the students to know that, okay, this is infographics, design is important, so therefore this one is 70 per cent, for example, just giving a number. As in, like, you're clearing doubts in our mind, okay, so the design is only 20 per cent; why is it like that? I guess it can help, but it will be tedious, so is it worth it? Probably?" (Student #1)

The results show that tertiary students prefer to be given time to digest the rubric and discuss it with the instructor. This is so that students' understanding of the rubric can be contextualised, and lecturers can make amendments to the rubric based on the students' comments if necessary. This finding is in line with the study done by Kite and Phongsavan (2017), Matshedisho (2020), and Mok and Toh (2015), which concluded that educators should communicate with the learners about the rubrics to narrow the gap in expectations between the two parties. Furthermore, a few respondents also suggested that students can better understand the rubric when educators provide feedback according to it.

"The lecturer can give us feedback based on the rubrics so that we can improve and revise it based on the rubrics. Because we might not be sure whether we answer the question correctly even though we follow the rubrics." (Student #5)

"Ya, when we get our result, we could see, for example, based on the rubrics, this part I get 15 marks out of 20 marks, that part I get five marks out of 20 marks. That is how we can understand the rubrics." (Student #3)

Feedback explicitly linked to rubric criteria allows students to identify the areas for improvement and improve their use of assessment rubrics in the future (Kite & Phongsavan, 2017; Venning & Buisman-Pijlman, 2013). Essentially, merely designing and giving students rubrics is inadequate, meaningful conversations, reflection and adjustment need to take place to ensure quality implementation of RRA (Kite & Phongsavan, 2017; Panadero & Jonsson, 2020; Sadler, 2014; Venning & Buisman-Pijlman, 2013).

Based on the theoretical framework, these findings highlight the potential solutions for resolving the aforementioned contradictions in the activity system. Training instructors to apply rubrics more consistently, allowing time for students to process the rubrics, and having two-way communication between instructor and students about the rubrics could foster a shared understanding among the instructors as well as the students (subjects). This allows the rubrics (mediating tools) to function more uniformly in the activity system and enables students to engage with the assessments (object) more effectively. This could potentially enhance students' overall experience with RRA (outcome).

Customising Rubric Design for Inclusive Assessment

The students suggested that assignments should allow for diverse perspectives, which would require rubrics to be more inclusive.

"There was one semester that the question asked us for an opinion from the perspective of a teacher, then the rubrics were not suitable for Student #3." (Student #5)

"I need to request back from Prof., can I answer from an army's perspective because I really do not know how to answer from a teachers' perspective..." (Student #3)

"If Prof. were strict to the rubrics, he would have a disadvantage." (Student #5).

"For example, in one of the assignments by Prof., she opened the question for views from different fields, so I can give my answer from an engineer's point of view, and he can give an answer from an army's point of view." (Student #2) The results indicate that by incorporating diverse perspectives, rubrics can become more inclusive, catering to a wider range of student experiences and backgrounds. Sadler (2014) mentioned a similar idea, stating that valid students' responses start from quality assessment tasks. In terms of rubrics' content, a respondent mentioned that learners at different stages should be provided with different rubrics.

"For me, I think the description of the rubric should cater to different levels of learners. For example, for younger learners like primary school students, tale writing essays, for example, it will include content like grammar and others. But for adults, it is more towards idea, more towards opinion, right." (Student #2)

These comments show that varying levels of learners should have individualised rubrics. This also suggests that learners with particular distinctive backgrounds may be more vulnerable to the challenges of RRA. This comment is congruent with the findings by Shadle et al. (2012), which emphasise the customisation of rubrics to accommodate different learners' backgrounds for a more accurate representation of their achievement levels. As they were discussing the rubric's design, another student pointed out that she would add specific numbers in the description of the rubrics, and another student gave his comment, too.

"One more thing is, if I were to develop a rubric, I would add in numbers, as in how many main points. Like for example, if the student wants to get full marks, "write five main points" or something like that. more specific. Like the second-best category, maybe it's 4 points. Then the student can be like, okay, I want the highest mark, so I will research 5 points and elaborate on them." (Student #5)

"I think rubrics have to go for quality and quantity. For example, if the student gives five main points, good the target is there, but if the five main points are not related to the question or topic, then ah, marks will be deducted." (Student #3)

The respondents who previously voiced concerns about the abstract subtleties in the rating scale description suggested that a holistic rubric might be more straightforward.

"Instead of putting like 1–5 marks, 16– 20 marks, the lecturer can just put the description for one full mark category." (Student #3)

"Holistic." (Student #5)

"You do not need the lower marks description because this is like, oh, if I am a lazy student, I only go for the lower marks criteria. But are there students who want low marks for their results? Of course, we all want to have high marks right." (Student #3)

The respondents proposed the idea of including specific numbers in the description of the rubric or simplifying the rubric by having a description for only one highest rating. According to Brookhart (2013), criteria can be described quantitatively, but numbers must be considered a guide, not a hard rule. It also comes down to the type of task because some assessments of creative work may not be suitable for specifying numbers. Brookhart (2013) also stated that it is crucial to have a description of the lowest to the highest rating so that it is accessible for students from a spectrum of performance, helpful for identifying areas for improvement and giving detailed feedback.

From the perspective of AT, the use of inclusive rubrics, holistic rubrics, and the inclusion of specific numbers in rubrics could align the rubrics (mediating tools) more closely with the needs of the students (subject), allowing them to engage with the assessments (object) with clearer expectations. These amendments enable rubrics to be a more effective mediating tool, which could improve students' experience with RRA (outcome). The summary of the themes and categories derived from the data is illustrated in Table 3.

CONCLUSION

This study aimed to explore the rubric's limitations and areas of consideration from the postgraduate students' angle. As illustrated by the AT (Vygotsky, 1978), different students (subjects) interact differently with assessment rubrics (mediating tools) and academic assessments (object), which led to different experiences with RRA, including the negative experiences discussed in this

Research objectives	Themes	Categories
1. To explore postgraduate students' perspectives on the limitations of RRA.	Impact of standardisation on autonomy among students and educators	Restriction of students' self-expressionLack of flexibility in adapting to diverse needs
	Challenges in fair and consistent assessment	Inconsistency in scoring across lecturersConfusion among students due to abstract nuances in descriptions
2. To explore postgraduate students' perspectives on the areas of consideration in improving RRA.	Collaborative and feedback-driven rubric implementation	 Enhancing communication and iterative adjustment Providing feedback aligned with rubric criteria
	Customising rubric design for inclusive assessment	 Incorporating diverse perspectives in assignments Differentiating rubrics for different learning levels Integrating quantitative and qualitative criteria Considering holistic rubric approaches

 Table 3

 Themes and categories derived from the data

study. From the diverse experiences, the participants suggested potential remedies to overcome the shortcomings of RRA in education.

In general, the respondents shared that assessment rubrics have the risks of limiting students' creative self-expression, invoking confusion among the students, and contradicting learners' needs. While students perceived that rubrics might create inconsistency in scoring when utilised by numerous lecturers, learners need to be able to discern between similarity and consistency. Further, students must understand that the intent of implementing RRA is not to achieve identical learning outcomes but rather to ensure consistent and transparent grading.

With regards to the suggested ways of improvements, the central idea was the intention to receive feedback from the users of the rubric and the flexibility of rubrics to be revised when necessary. The other recommendations consist of ensuring the inclusivity of assessment, having different rubrics for different learners, including both quantitative and qualitative content in the description of criteria, having rubrics with only one rating category, and providing feedback to students based on the rubric. The inputs given by the respondents were in parallel with the three features of rubrics mentioned in the literature review, which are specificity, inter-rater reliability, and flexibility. Hence, this study signifies that rubrics have a few weaknesses despite the many benefits proclaimed in various previous studies. Rubrics also have several potential remedies that should be tested for validation. In short, an effective incorporation of RRA is tedious but necessary so that learners can have a quality education.

Implications of the Study

The Program Learning Outcome (PLO) of the master's program in this study encompasses developing professionals who are knowledgeable and highly skilled in educating and leading autonomously and developing educational technology through sustainable teaching innovation. The Course Learning Outcome (CLO) includes demonstrating leadership qualities, conducting research, and using digital tools to resolve problems in the field of educational technology. In relation to the mentioned PLO and CLO, this study highlights the need for rubrics to evaluate theoretical knowledge and practical skills, such as students' ability to lead projects, the rigour of their research, and proficiency in using digital tools to solve real-world problems. This necessitates the development of dynamic and adaptable rubrics with criteria that capture the mentioned competencies' nuances.

Through understanding the students' experience, the findings suggested that communication and flexibility are essential in the implementation of rubric-referenced assessments. Specifically, for novice users of rubrics, lecturers should give students time to digest the rubric and discuss it for clarification (Kite & Phongsavan, 2017; Matshedisho, 2020). Moreover, the findings from this research also suggest the necessity for fostering collaboration among faculty members (Postmes et al., 2023). If the rubric is intended for use across different lecturers, conducting rubric review sessions and providing professional development on its implementation become imperative among the teaching staff. A scoring calibration session whereby a group of lecturers score a set of student work and discuss their rationale can help identify the discrepancies in interpretation and establish a common understanding of the rubric (Sadler, 2014; Venning & Buisman-Pijlman, 2013). After consolidating feedback via different forms of communication, the rubric should be flexible enough to be revised to ensure the clarity and consistency of the rubric's criteria.

Based on the discoveries of this study, educators are prompted to focus on emphasising the diversity of assessment before delving into rubric design (Sadler, 2014). This would allow all students to articulate their ideas creatively (Bennett, 2016). Given the evident complexity of designing and implementing RRA, this study underscores the need for policy formulation. Institutional leaders and decision-makers in institutions play a significant role in encouraging educators to integrate rubrics into their instruction. For instance, for lecturers who would implement assessment rubrics, more capacity in the working schedule should be given. A budget can be allocated to employing teaching assistants to assist in delegating the lecturers' workload, and comprehensive training ought to be introduced to raise awareness and proficiency among the lecturers concerning the effective utilisation of assessment rubrics.

Viewing the research findings through the lens of AT also provides several theoretical implications. As postulated by the AT, while rubrics (mediating tools) were intended to guide and support learning, they could impose external constraints on student learning outcomes (outcome), such as limiting students' creative expression in completing their assessments (object; Vygotsky, 1978). This raises a dilemma about the balance between providing structure and fostering individual agency within a learning environment (Sitorus, 2020). AT focuses on the pertinence of social interaction and collaboration in knowledge construction. The limitations of rubrics could be addressed by collaborative rubric development and ongoing communication among both instructors and learners, building a shared understanding within the learning community (Kite & Phongsavan, 2017; Panadero & Jonsson, 2020; Sadler, 2014; Venning & Buisman-Pijlman, 2013).

Further, AT assumes that knowledge is constantly evolving and mediated through interactions, which highlights the risks of a rigid rubric (Engeström, 1993). This indicates that assessment tools should be adaptable and fluid, allowing for ongoing feedback and adjustments based on specific learning and learner needs (Sadler, 2014). In short, viewing the limitations and areas for consideration of rubrics through the lens of AT offers valuable insights into the complex interplay between subject, tools, object, and outcome. These implications can guide practitioners in developing more effective assessment practices.

Limitations of the Study

This study has postulated several key findings on RRA, but a few limitations

exist. Firstly, the respondents of this study comprised a small group of postgraduate students and the research outcomes were based on their self-reported data. Thus, the results of this study cannot be generalised to populations of students with different demographics from the respondents in this study. In qualitative research, researchers are the main instrument in collecting data, and the authors acknowledged that the findings in this study are prone to the researchers' bias and idiosyncrasies. It is recommended that quantitative research on relevant topics be done with a larger sample size to validate the findings of this study. Furthermore, this study explored the limitations and areas of consideration for analytic rubrics only; similar research can be conducted on other types of rubrics, such as holistic rubrics, in the future.

Recommendations for Future Studies

The findings and insights from this study further call for future research endeavours. enriching the understanding of RRA and its implications. Firstly, while this study focused on postgraduate students' perspectives, future research could incorporate educators' viewpoints on RRA. Exploring educators' experiences with RRA and how they perceive the challenges students face when implementing RRA could lead to a more holistic understanding. As this study explored a few areas of refining the design and application of RRA, future studies can delve into effective strategies for integrating RRA in a way that addresses the identified limitations, optimising the implementation of RRA for future practitioners. Moreover, mixed-methods approaches can be employed in future research. Combining qualitative and quantitative methods can provide a deeper understanding of the identified pitfalls and areas of improvement. In future research, it may also be beneficial to extend the application of a theoretical framework beyond AT (Vygotsky, 1978) to encompass Engeström's (1999) Cultural Historical Activity Theory. By incorporating social, cultural, and historical factors, researchers could develop an RRA that is methodologically sound, culturally sensitive and contextually appropriate, leading to more equitable and effective assessment practices in higher education.

ACKNOWLEDGEMENT

The authors would like to thank the Research Management Centre and Faculty of Educational Studies, Universiti Putra Malaysia, for their support in the publication of this manuscript. The authors are also grateful to the Centre for Academic Development and Leadership Excellence (CADe-Lead) Universiti Putra Malaysia for the *Geran Insentif Penyelidikan dalam Pengajaran dan Pembelajaran* (GIPP: 9323783, August 2021-January 2024) which made this research possible.

REFERENCES

Anandi, R. P., & Zailaini, M. A. (2019). Using Rasch Model to assess self-assessment speaking skill rubric for non-native Arabic language speakers. *Pertanika Journal of Social Sciences* & Humanities, 27(3), 1469-1480.

- Andrade, H. L., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation, 10*(1) Article 3. https://doi.org/10.7275/g367-ye94
- Bearman, M., & Ajjawi, R. (2021). Can a rubric do more than be transparent? Invitation as a new metaphor for assessment criteria. *Studies in Higher Education*, 46(2), 359-368. https://doi. org/10.1080/03075079.2019.1637842
- Bennett, C. (2016). Assessment rubrics: Thinking inside the boxes. *Learning and Teaching*, 9(1), 50-72. https://doi.org/10.3167/latiss.2016.090104
- Brinkmann, S. (2014). Unstructured and semistructured interviewing. In P. Leavy (Ed.), *The* Oxford Handbook of Qualitative Research (pp. 277-299). https://doi.org/10.1093/ oxfordhb/9780199811755.013.030
- Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading.* ASCD.
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3, Article 22. https://doi.org/10.3389/ feduc.2018.00022
- Bukhari, N., Jamal, J., Ismail, A., & Shamsuddin, J. (2021). Assessment rubric for research report writing: A tool for supervision. *Malaysian Journal of Learning and Instruction*, 18(2), 1-43. https://doi.org/10.32890/mjli2021.18.2.1
- Chowdhury, F. (2019). Application of rubrics in the classroom: A vital tool for improvement in assessment, feedback and learning. *International education studies*, *12*(1), 61-68. https://doi. org/10.5539/ies.v12n1p61
- Cockett, A., & Jackson, C. (2018). The use of assessment rubrics to enhance feedback in higher education: An integrative literature review. *Nurse Education Today*, 69, 8-13. https://doi. org/10.1016/j.nedt.2018.06.022

- Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). Sage Publication.
- Darus, S., Stapa, S. H., & Hussin, S. (2003). Experimenting a computer-based essay marking system at Universiti Kebangsaan Malaysia. Jurnal Teknologi, 39(1), 1-18. https://doi. org/10.11113/jt.v39.472
- Engeström, Y. (1993). Developmental studies of work as a testbench of activity theory: The case of primary care medical practice. In S. Chaiklin & J. Lave (Eds.), Understanding practice: Perspectives on activity and context (pp. 64-103). Cambridge University Press. https://doi. org/10.1017/CBO9780511625510.004
- Engeström, Y. (1999). Innovative learning in work teams: Analyzing cycles of knowledge creation in practice. In Y. Engeström, R. Miettinen, & R.-L. Punamäki (Eds.), *Perspectives on activity theory* (pp. 377-406). Cambridge University Press.
- Fraile, J., Panadero, E., & Pardo, R. (2017). Cocreating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation*, 53, 69-76. https://doi.org/10.1016/j.stueduc.2017.03.003
- Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? Operationalising data saturation for theory-based interview studies. *Psychology* and Health, 25(10), 1229-1245. https://doi. org/10.1080/08870440903194015
- Hamilton, H., Gurak, E., Findlater, L., & Olive, W. (2000). *The confusion matrix*. https://www2. cs.uregina.ca/~hamilton/courses/831/notes/ confusion matrix/confusion matrix.html
- Holmstedt, P., Jönsson, A., & Aspelin, J. (2018). Learning to see new things: Using criteria to support pre-service teachers' discernment in the

context of teachers' relational work. *Frontiers in Education, 3*, Article 54. https://doi.org/10.3389/feduc.2018.00054

- Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. Assessment & Evaluation in Higher Education, 39(7), 840-852. https://doi.org/10.1080/02602938.2013.875117
- Kite, J., & Phongsavan, P. (2017). Evaluating standards-based assessment rubrics in a postgraduate public health subject. Assessment & Evaluation in Higher Education, 42(6), 837-849. https://doi.org/10.1080/02602938.2016.1199773
- Krueger, R. A. (2014). *Focus groups: A practical guide for applied research*. Sage publications.
- Marques, J. F., & McCall, C. (2005). The application of interrater reliability as a solidification instrument in a phenomenological study. *The Qualitative Report*, 10(3), 439-462.
- Matshedisho, K. R. (2020). Straddling rows and columns: Students' (mis)conceptions of an assessment rubric. Assessment & Evaluation in Higher Education, 45(2), 169-179. https://doi.or g/10.1080/02602938.2019.1616671
- Ministry of Higher Education Malaysia. (2021). Alternative Assessment in Higher Education: A Practical Guide to Assessing Learning. Department of Higher Education Malaysia. https://utmcdex.utm.my/wp-content/ uploads/2024/09/EBOOK-Alternative-Assessment-in-Higher-Education-2022.pdf
- Mok, J. C. H., & Toh, A. A. L. (2015). Improving the ability of qualitative assessments to discriminate student achievement levels. *Journal* of International Education in Business, 8(1), 49-58. https://doi.org/10.1108/JIEB-12-2013-0048
- Noh, N. M., Zain, M. R. M., Hamid, Y. S., Bakar, I. A. A., & Mohamad, M. (2021). Analytic rubric in evaluating the continuous assessment in projects of civil engineering undergraduate students in dynamics subject. *Asian Journal of*

Liang Jing Teh, Su Luan Wong, Mas Nida Md Khambari, Rahmita Wirza O. K. Rahmat and Sai Hong Tang

University Education, *17*(4), 352-366. https:// doi.org/10.24191/ajue.v17i4.16221

- Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal* of the American Society for Information Science and Technology, 60(5), 969-983. https://doi. org/10.1002/asi.21030
- Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, 30, Article 100329. 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/10.108 0/0969594X.2013.877872
- Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage Publications.
- Pérez-Guillén, S., Carrasco-Uribarren, A., Celis, C. L., González-Rueda, V., Rodríguez-Rubio, P. R., & Cabanillas-Barea, S. (2022). Students' perceptions, engagement and satisfaction with the use of an e-rubric for the assessment of manual skills in physiotherapy. *BMC Medical Education*, 22, Article 623. https://doi.org/10.1186/s12909-022-03651-w
- Popham, W. J. (1997). What's wrong--and what's right--with rubrics. *Educational leadership*, 55(2), 72-75.
- Postmes, L., Bouwmeester, R., de Kleijn, R., & van der Schaaf, M. (2023). Supervisors' untrained postgraduate rubric use for formative and summative purposes. Assessment & Evaluation

in Higher Education, *48*(1), 41-55. https://doi.or g/10.1080/02602938.2021.2021390

- 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
- Saeed, K. M., Ismail, S. A. M. M., & Eng, L. S. (2019). Malaysian speaking proficiency assessment effectiveness for undergraduates suffering from minimal descriptors. *International Journal* of Instruction, 12(1), 1059-1076. https://doi. org/10.29333/iji.2019.12168a
- Shadle, S. E., Brown, E. C., Towns, M. H., & Warner, D. L. (2012). A rubric for assessing students' experimental problem-solving ability. *Journal* of Chemical Education, 89(3), 319-325. https:// doi.org/10.1021/ed2000704
- Sitorus, M. L. (2020). Non-native English teachers interpretation of rubrics used for assessing students' writing. *Proceedings of the International Conference on Future of Education*, 3(2), 16-25. https://doi.org/10.17501/26307413.2020.3202
- Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In C. Willig & W. Stainton-Rogers (Eds.), *The Sage handbook of qualitative research in psychology* (2nd ed., pp. 17-37). Sage.
- Venning, J., & Buisman-Pijlman, F. (2013). Integrating assessment matrices in feedback loops to promote research skill development in postgraduate research projects. Assessment & Evaluation in Higher Education, 38(5), 567-579. https://doi.org/10.1080/02602938.2012.661842
- Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Harvard University Press.