

Investigating the Interrelationship between Teachers' Grading Practices and Classroom-Based Assessment Strategies among Malaysian Preservice Teachers

Shahazwan Mat Yusoff^{1*}, Anwar Farhan Mohamad Marzaini², Ahmad Muhaimin Mat Jusoh³, Siti Nadya Zynuddin³, Mohd Helme Basal⁴, Noorhayati Zakaria⁵

¹Department of Curriculum & Instructional Technology, Faculty of Education, Universiti Malaya, 50603 Kuala Lumpur, Malaysia.
shahazwan@um.edu.my

²Academy of Language Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, 14000 Permatang Pauh, Pulau Pinang, Malaysia.
anwarmarzaini@uitm.edu.my

³Department of Educational Management, Planning, and Policy, Faculty of Education, Universiti Malaya, 50603 Kuala Lumpur, Malaysia.
ahmadmuhaimin@um.edu.my
nadya.zynuddin@um.edu.my

⁴Faculty of Sports Science & Recreation, Bangunan Akademik 3, Universiti Teknologi MARA, 40450 Shah Alam, Selangor.

⁵Department of Language and Literacy Education, Faculty of Education, Universiti Malaya, 50603 Kuala Lumpur, Malaysia.
noorhayati1205@gmail.com
*Corresponding author

<https://doi.org/10.24191/ajue.v21i1.5622>

Received: 27 September 2024

Accepted: 27 January 2025

Date Published Online: 11 March 2025

Published: 11 March 2025

Abstract: Within the Malaysian educational landscape, the grading practices and classroom-based assessment strategies adopted by preservice school teachers have become a crucial area of investigation as they may significantly impact students' academic achievements and overall learning experiences. The primary aim of this study is to explore the relationship between teachers' grading practices (TGP) and CBA practices among preservice school teachers. TGP encompasses five dimensions: Importance (IMPC), Student Effort (STUE), Student Ability (STUAB), Usefulness (USE), and Teachers' Grading Habits (TGH). A quantitative research design, specifically a survey-based approach, was employed in this study. The instrument was developed and validated through face and content validity, a pilot study, and Exploratory Factor Analysis (EFA). The quantitative data were primarily analysed using Partial Least Square Structural Equation Modelling (PLS-SEM). A total of 338 preservice teachers from 4 public universities in Selangor participated in the study, completing a 5-point Likert-scale questionnaire. Findings indicate that among the five dimensions, IMPC, STUE, and TGH significantly influence CBA practices, whereas STUAB and USE do not. The results suggest that preservice teachers prioritise student effort and well-established grading habits in their assessment strategies, reinforcing the importance of structured and transparent grading approaches. However, the limited impact of student ability and perceived usefulness on CBA implementation highlights the challenges of aligning grading practices with educational assessment frameworks. This study contributes to the understanding of grading practices in Malaysian teacher education by identifying key factors that shape assessment decisions. The findings highlight the need for continuous refinement of assessment strategies to ensure consistency and fairness in student evaluations.

Keywords: Teachers' Grading Practices, Classroom-based Assessment, Assessment Practices

1. Introduction

The current grading practices among preservice school teachers involve a multifaceted approach considering academic achievement and factors such as student effort and work habits. This complex decision-making process, documented in several studies (Guskey, 2011; Randall & Engelhard, 2009; Yesbeck, 2011), needs to be revised to align with the approach recommended by educational measurement experts. These experts argue that grading should be strictly based on academic achievement or mastery of learning standards, excluding confounding factors like effort and work habits (Dyrness & Dyrness, 2008; McMillan, 2008; Reynolds, 1989; O'Connor, 2007; Wormeli, 2006). This divergence is identified by Brookhart (1991, 1993) as a validity issue best framed within Messick's (1989) validity framework. This framework compels us to explore how preservice teachers understand grades, the value judgments they make when grading, how they interpret what a grade represents, their perspectives on the use and consequences of grades, and their significance to grades. This presents a substantial issue since preservice teachers' grades can significantly affect classroom-based assessment and student learning, engagement, and motivation.

However, this validity problem in preservice teachers' grading practices needs to be further examined, particularly in the context of preservice education. Given that grades play a pivotal role in students' educational journey, influencing their progression, self-esteem, and future educational opportunities, it is crucial to delve deeper into preservice teachers' grading practices and their impact on classroom-based assessment. A comprehensive understanding of these practices can offer valuable insights to refine assessment procedures, enhance fairness, and foster an optimal learning environment in preservice classrooms. Furthermore, it can help bridge the gap between the actual grading practices and the recommended practices by the measurement community, enhancing the validity of the grading process.

Despite the emerging recognition of Classroom-Based Assessment (CBA) in Malaysia, its implementation is often challenged by an examination-oriented assessment system still prevailing in the educational structure (Fadila et al., 2016). Yuh and Kenayathulla (2020) articulate that the dichotomy between traditional examination-based assessments and CBA creates difficulties for preservice teachers to integrate CBA into their daily pedagogical practices. This disparity in understanding and application of CBA could distort the grading process and undermine the intrinsic values of CBA (Panadero & Jonsson, 2013). However, Varatharaj et al. (2015) suggest ample room for improving grading practices within CBA, assuming that preservice teachers recalibrate their assessment strategies. Abd Mutalib and Ahmad (2012) highlight those challenges in CBA implementation stem from deficiencies in Pedagogical Content Knowledge, biased decision-making, and lesson planning insufficiencies. Such shortcomings can lead to underestimating students' abilities, compromising performance reporting accuracy.

Although extensive studies have been conducted on School-Based Assessment (SBA) within schools in Malaysia (Bidin & Mahamod, 2016; Ambotang & Gobalakrishnan, 2017; Moktar et al., 2018; Isa & Naim, 2013; Mazlini Adnan & Noorfazelawati Abd Kadir, 2014; Che Aleha Ladin & Md Nasir Ibrahim, 2015; Ahmad & Mahamod, 2016), research focusing on the implementation of CBA among preservice teachers in local settings remains scant. Addressing this knowledge gap is paramount, particularly in providing a holistic understanding of assessment practices in the context of CBA and exploring preservice teachers' perspectives toward these practices. Thus, the principal aim of this study is to investigate the impact of preservice teachers' grading practices on the implementation of CBA, thereby offering invaluable insights to educators and policymakers for enhancing the quality and effectiveness of classroom-based assessments.

2. Literature Review

2.1 Teacher Grading Practices (TGP)

The efficacy of classroom assessment and grading practices has increasingly become a pivotal topic of exploration in education (McMillan & Lawson, 2001; McMillan et al., 2002). Liu (2008) posits that classroom assessment and grading practices are inherently intertwined. As Liu (2008) explicates, grading represents accountability for both schools and students, with significant implications for

instruction and curriculum. Furthermore, grading is instrumental in assessing teaching effectiveness, facilitating informed decisions regarding students' progress and the efficacy of pedagogical approaches. Furthermore, grades can provide students with insightful feedback on their areas of strength and aspects that require improvement (Shepard, 2017; Rahman, 2020).

Educators, especially those from non-traditional or alternative teaching certification programs, need more training in practical assessment or grading techniques, as Smith (2014) and Linton (2017) suggested. With alternative certification programs gaining popularity, there is an increasing need to examine the differences in teacher training between traditional and non-traditional certification programs, as noted by Redding and Smith (2016) and Abdul Razak et al. (2023). Moreover, the lack of pedagogical training among these teachers and their consequent use of ineffective classroom methods that lead to students' discouragement and disengagement has been emphasised in recent educational reforms, making it crucial to investigate the impact of non-traditionally trained teachers in the classroom, according to Brewer and deMarrais (2015) and Mohamad Marzaini et al. (2024). Previous research on grading has often focused on examining past trends and patterns that reinforce ineffective grading and assessment practices in the classroom (Mohamad Marzaini et al., 2024; McMillan et al., 2002; O'Connor, 2007; Reeves, 2011). According to Brookhart et al. (2016), numerous studies have found that grades are typically influenced by factors that teachers deem essential, such as effort, ability, work habits, participation, attendance, and others, which can vary greatly. This lack of consistency in grading criteria limits the effectiveness of grades in conveying specific levels of academic mastery to students, parents, and other stakeholders involved in the learning process (Abdul Razak et al., 2023; Link, 2018).

According to Pirzad and Abadikhah (2022), teachers are grade developers. Outside of Malaysia, teachers in countries such as Brazil (Zapata et al., 2015), Finland Kaplan et al. (2020), and the Czech Republic Vykyddal et al. (2020) produce marks or grades based on their assessments. When teachers are in charge of grading, they are often required to make decisions based on accomplishment variables. For example, Kubiszyn and Borich (2016, p.24) emphasised that grades should only reflect academic achievement. Brookhart and Nitko (2019) defined "grading variables" as indicators of content or standard performance that were addressed independently from measures that reflect effort or social conduct.

Teachers utilise multiple factors when assigning grades, including student attitude, engagement, interest, effort, classroom behaviour, and progress (Mohamad Marzaini et al., 2024; Russell & Austin, 2010; Willingham et al., 2002; Randall & Engelhard, 2010; Kunnath, 2017). According to Green et al. (2007), most teachers and teacher candidates believe it is ethical to incorporate growth, effort, class participation, attendance, classroom behaviour, and work habits when determining grades. While teachers may be constrained by formal grading policies set by their school or district, they still consider various constructs when evaluating student performance.

While various initiatives to reform grading practices have been instituted, effecting substantial changes in the actual grading practices of educators might pose substantial challenges due to their prevailing beliefs and perceptions (Mat Yusoff et al., 2025; Olsen & Buchanan, 2019). For instance, in the Canadian context, despite introducing standardised reporting protocols, educators remained predominantly supportive of improvement-based grading methodologies (Tierney et al., 2011). In Australian educational settings, Wyatt-Smith et al. (2010) illustrated that educators' evaluative decisions were often premised on the attributes and activities of the students rather than the evidential bases of assessment, even within a standards-based paradigm. Further, a study conducted by Olsen and Buchanan (2019) revealed that educators in the United States exhibited limited adherence to standardised grading guidelines. They frequently based their grades on non-standardized student work samples and relied on individualised criteria for evaluating student work.

2.2 Classroom-based Assessment in Malaysia

The implementation of CBA in Malaysia represents a critical shift from traditional examination-based evaluation towards a more holistic approach to assessing student learning. This change aligns with the objectives outlined in the Malaysian Education Blueprint 2013-2025, which emphasises fostering higher-order thinking skills and a more comprehensive understanding of students' progress (Ministry of Education Malaysia, 2013). Unlike standardised tests that often rely on closed-

ended questions and psychometric techniques, CBA employs a range of assessment methods, including both traditional and alternative approaches, to capture a wider array of student abilities and achievements. Traditional assessments, such as multiple choice or true/false questions, are complemented by performance-based evaluations, such as oral presentations and hands-on projects, to provide a more thorough evaluation of a student's knowledge and skills (Ansori et al., 2019; Collins, 2014).

Despite the potential benefits, the implementation of CBA in Malaysian primary schools faces several challenges. A significant barrier is the variability in teachers' understanding and application of assessment strategies. Many teachers lack confidence in using alternative assessment methods and need more training effectively to integrate these methods into their teaching practices. This inconsistency can lead to variations in assessment quality and effectiveness across different schools (Mohamad Marzaini et al., 2024; Mansor et al., 2019; Chappuis et al., 2020). Additionally, the shift towards CBA requires a substantial change in the roles of teachers, who must now act as both facilitators and assessors. This dual role demands a deep understanding of formative assessment practices to support student learning while simultaneously gathering accurate data on student progress (Abdul Razak et al., 2023; Black & Wiliam, 1998; Stiggins et al., 2005).

The Malaysian Ministry of Education has attempted to address these challenges by issuing guidelines, such as the "Panduan Pelaksanaan Pentaksiran Bilik Darjah" (Classroom Assessment Implementation Guide), aimed at standardising CBA practices across schools (Kementerian Pendidikan Malaysia, 2019). However, ensuring consistency and fairness in assessment remains a challenge due to differing levels of teacher readiness and resource availability. Research suggests that more extensive professional development and support are needed to build teachers' capacity to implement CBA effectively (Mat Yusoff et al., 2025; Faizah, 2011; Narinasamy & Nordin, 2018).

Therefore, while the adoption of CBA in Malaysian schools is generally viewed positively for its potential to reduce exam pressure and provide more personalised feedback, its successful implementation is contingent upon overcoming challenges related to teacher training, resource allocation, and maintaining consistency in assessment practices. Addressing these challenges will require continuous policy efforts and a commitment to professional development to ensure that CBA can fully realise its potential to enhance student learning and development (Lewkowicz & Leung, 2020; Lokman, et al., 2024).

2.3 Grading Practices and Classroom-based Assessment

Research has also revealed that teachers who utilise effective grading methods have more favourable attitudes about assessment and are more likely to believe they possess the knowledge and abilities required to use assessment successfully (Shepard et al., 2018). Pinger et al. (2018) found that teachers who held more constructivist views of grading, which emphasise the use of multiple sources of evidence and the development of student self-assessment skills, were more likely to use formative assessment and provide detailed, specific feedback to students. This can be perceived that teachers who understand constructivist grading theories are more likely to use assessment as a tool for promoting student learning and development. Constructivist grading theories place a strong emphasis on the use of formative assessments to learn more about student understanding and development (Mat Yusoff et al., 2025; Shepard et al., 2018). This knowledge is then used to inform instruction and give students who need more specialised support (Lijie et al., 2024; Horsburgh & Ippolito, 2018). The teachers who see that assessment should be used to track and aid in the dynamic process of student learning also understand how crucial it is to give students chances to evaluate their learning and get constructive criticism on how they are doing (Pitt & Carless, 2022). By allowing teachers to pinpoint areas where students need extra assistance and modify lessons as necessary to enhance student learning and growth, this method can improve student outcomes.

Preservice teachers who utilise effective grading methods often hold more favourable attitudes about assessment and are more likely to believe they possess the knowledge and abilities required to use assessment successfully (Mat Yusoff et al., 2024; Shepard et al., 2018). Pinger et al. (2018) found that preservice teachers who held more constructivist views of grading, which emphasise the use of multiple sources of evidence and the development of student self-assessment skills, were more likely to use formative assessment and provide detailed, specific feedback to students. This suggests that

preservice teachers who understand constructivist grading theories are more likely to use assessment as a tool for promoting student learning and development. Constructivist grading theories place a strong emphasis on the use of formative assessments to learn more about student understanding and development (Mohamad Marzaini et al., 2024; Shepard et al., 2018). This knowledge is then used to inform instruction and provide specialised support to students who need it (Mat Yusoff et al., 2025; Horsburgh & Ippolito, 2018). Preservice teachers who perceive assessment as a means to track and aid in the dynamic process of student learning understand the importance of giving students opportunities to evaluate their learning and receive constructive feedback on their progress (Pitt & Carless, 2022). By allowing preservice teachers to pinpoint areas where students need extra assistance and modify lessons as necessary to enhance student learning and growth, this method can improve student outcomes.

Feldman (2018) illustrated that preservice teachers who held more traditional views of grading, which emphasised the use of standardised tests and objective measures, were more likely to use grades as a primary means of communicating student performance while providing less detailed feedback. This suggests that preservice teachers who understand traditional grading theories may be less likely to use assessment as a tool for promoting student learning and development. Traditional grading theories frequently emphasise summative assessments to evaluate student performance and assign grades, often administered after a unit or term to measure overall comprehension. Such models may focus less on formative assessments for tracking student knowledge and guiding instruction (Abdul Razak., 2023; Shepard, 2019; Kulasegaram & Rangachari, 2018). However, many preservice teachers use a mixture of formative and summative assessments, as both are crucial for assessing student understanding and progress (Mohamad Marzaini et al., 2024; Cizek et al., 2019).

Additionally, research carried out by Alt (2018) has shown that preservice teachers with constructivist views of grading are more likely to use formative assessment and provide detailed, specific feedback. This highlights the importance of preservice teachers understanding constructivist grading theories to use assessment effectively as a tool for student learning and development. Constructivist grading theories emphasise formative assessments to better understand student learning and use that knowledge to inform teaching and provide needed support (Mohamad Marzaini et al., 2024; Matthewman et al., 2018). By focusing on student needs and allowing them to evaluate their progress, preservice teachers can adapt lessons to improve outcomes. This approach helps preservice teachers use assessment data to enhance student learning and development (Mat Yusoff et al., 2024; Shepard et al., 2018).

On the other hand, research by Gentrup et al. (2020) found that preservice teachers employing norm-referenced grading where students' performance is compared against their peers tend to rely heavily on grades to communicate progress and provide less in-depth feedback. Norm-referenced grading methods focus on comparing student performance to a class or grade-level average (Abdul Razak et al., 2023; Gaertner, 2022). While this approach is valuable for summative evaluations, it may not emphasise formative assessments that inform instruction or provide tailored support for students. Preservice teachers combining norm-referenced and criterion-referenced grading methods may still incorporate formative and summative assessments to balance learning support and progress evaluation.

2.4 The Interplay of Usefulness and Importance in Preservice Teachers' Grading Practices within Classroom-Based Assessments

Preservice teachers' perceptions of the usefulness and importance of grading are closely intertwined and significantly influence their assessment practices in the context of CBA. Preservice teachers who view grading as highly useful tend to see it as an integral tool for guiding instructional strategies and providing constructive feedback that promotes student learning and development. This perception encourages them to adopt formative assessment methods, such as ongoing feedback, which engage students actively in their learning journey and help them understand their progress and areas for improvement (Bürgermeister et al., 2021; Coombs et al., 2018). In this way, the perceived usefulness of grading fosters a learning environment where assessment is seen not just as a final judgment but as a continuous process that supports and enhances learning.

Simultaneously, the importance of grading, as perceived by preservice teachers, plays a crucial role in how they implement assessment practices. Preservice teachers who consider grading to be of high importance often emphasise fairness and consistency, aligning their practices with clear learning

objectives and criteria. This alignment ensures that assessment contributes meaningfully to student motivation, guiding both learners and educators toward achieving set educational goals.

In the context of CBA, this perceived importance compels preservice teachers to focus on both the accuracy and the impact of their grading decisions, making sure that they are not only measuring academic performance but also fostering a positive learning environment that encourages effort and improvement (Mohamad Marzaini et al., 2024; Cheng et al., 2020; Coombs et al., 2018). By integrating both the usefulness and importance of grading, preservice teachers can develop a balanced approach to assessment that combines objective measures with continuous feedback, ultimately enhancing the learning experience for students and ensuring that grading practices serve as an effective tool for educational development (Coombs, Ge, & DeLuca, 2021; Cheng et al., 2020).

H1: Importance has a significant and positive influence on CBA

H2: Usefulness has a significant and positive influence on CBA

2.5 The Impact of Student Effort, Ability, and Teacher Habits on Grading Practices in Classroom-Based Assessments

Preservice teachers' grading practices within CBA frameworks are shaped by their perceptions of student effort, ability, and their grading habits, each playing a significant role in how assessments are conducted and interpreted. Student effort is often a critical consideration, as preservice teachers may view it as an indicator of a student's commitment and determination to learn (Getenet et al., 2024). By factoring effort into grading, preservice teachers aim to reward persistence and encourage a positive attitude toward learning. However, the inclusion of effort can introduce a level of subjectivity into the grading process, as the assessment of effort may vary significantly among preservice teachers, potentially impacting the fairness and reliability of grades. This subjectivity can lead to inconsistencies in how grades are awarded, as what one preservice teacher perceives as sufficient effort may differ from another's interpretation (Cheng, DeLuca, Braund, & Yan, 2018; McMillan, 2020; Getenet et al., 2024).

Student ability is another factor influencing grading practices where preservice teachers may modify grades based on their perception of a student's inherent or demonstrated abilities. Within CBA, preservice teachers may differentiate their grading to promote equity, ensuring that all students are encouraged to improve and achieve growth (Mahalingappa, 2024). While this approach can support a growth mindset and accommodate diverse learning needs, it can also lead to unintended consequences such as grade inflation or lower academic expectations for certain students. This practice may inadvertently undermine the goal of fair and objective assessment, highlighting the need for clearer guidelines, rubrics, and professional development to ensure consistency and maintain rigorous standards across different contexts (Alam & Mohanty, 2023; Brookhart, 2017).

Additionally, preservice teachers' grading habits, shaped by their beliefs, experiences, and contextual influences, play a crucial role in how they implement grading practices. Preservice teachers may develop habitual patterns, such as a tendency towards leniency or strictness, which can persist over time and affect the objectivity and fairness of assessments (Mahalingappa, 2024). These habits may be deeply rooted and resistant to change, even in the face of new educational policies or reforms aimed at improving assessment practices. For instance, preservice teachers who have become accustomed to particular grading methods may find it challenging to adapt to new guidelines that require more objective and standardised assessment criteria (Mat Yusoff et al., 2025; Urhahne & Wijnia, 2020; Guskey, 2019). This resistance underscores the importance of reflective practices that encourage preservice teachers to critically evaluate their grading habits and align them with contemporary educational objectives that promote fairness, transparency, and student development (Tai et al., 2023).

H3: Student Effort has a significant and positive influence on CBA

H4: Student Ability has a significant and positive influence on CBA

H5: Teachers' Grading Habits have a significant and positive influence on CBA

3. Methodology

3.1 Research Design

The primary aim of this quantitative study, employing a survey research design, is to evaluate teachers' grading practices and their classroom-based assessment practices. A quantitative approach was adopted to systematically examine the educational context, with a specific focus on teachers' perceptions, beliefs, and behaviours. Survey research, a widely recognised method in the social sciences, is particularly effective for systematically capturing information on human perceptions, beliefs, and behaviours (Singleton & Straits, 2009). This methodological choice aligns with the study's objectives, enabling the collection of data that can comprehensively address the research questions.

The quantitative approach facilitates the precise measurement of variables and the identification of relationships and patterns among them, thus providing a nuanced understanding of the factors influencing grading practices and classroom-based assessment practices. Through the utilisation of a survey research design, this study aims to uncover critical insights into the complexities of teachers' grading practices and assessment methodologies within the educational landscape. These insights have the potential to inform the design of effective teacher training programs and strategic interventions.

3.2 Sample

In this study, a simple random sampling method was employed to ensure a representative sample of preservice teachers from January 2024 to April 2024. The researchers ensured ethical standards by obtaining permission letters for distributing the instrument and obtaining approval from the University Malaya Research Ethics Committee (Non-medical) with an approval letter number UM.TNC 2/UMREC. Also, written consent was obtained from all the participants before they answered the questionnaire. As a widely used probability sampling technique, this sampling technique allows each member of the target population an equal chance of being selected, reducing selection bias and ensuring the generalizability of research findings to the larger population (Creswell, 2014; Miles & Huberman, 1994). The sample was drawn from 4 public universities in Selangor, comprising 338 preservice teachers, with a gender distribution of 73 males and 265 females. The participants' demographic information is detailed in Table 1. This diverse sample allows for a thorough investigation of grading strategies employed by preservice teachers with varying backgrounds and experiences.

Table 1

Demographic Information of the Participants

Gender	Frequency
Male	73
Female	265
Total	338
University	
University A	77
University B	89
University C	94
University D	78
Total	338

3.3 Instrumentation

In this study, a single survey instrument was utilised to collect data, focusing on classroom-based assessment (CBA) practices. The survey was adapted from Goggin (2018) with specific reference to classroom-based assessment practices, while the Teachers' Grading Practices (TGP) construct was derived from Link (2018). The development of the survey instrument was guided by "Classroom

Assessment for Student Learning" by Chappuis et al. (2020). To ensure face validity, consultations were held with five experts possessing extensive experience in classroom-based assessment. For content validity, the survey was disseminated to 20 additional experts in the field via email, ResearchGate, and social media platforms, including WhatsApp, Telegram, and Facebook Messenger. The selection of these experts was based on their research expertise, educational qualifications, and prior accomplishments in relevant research domains (Lynn, 1986). Out of the 20 experts approached, 10 agreed to participate, 2 declined due to time constraints, and the remaining did not respond. As recommended by Hong et al. (2019), incorporating input from experts with specialised knowledge in the research area enhances the content validity of the instrument and mitigates potential biases. The participating experts evaluated the instrument's relevance, clarity, and simplicity through a structured questionnaire. This rigorous validation process ensured that the survey instrument was both reliable and valid, thereby strengthening the methodological robustness and credibility of the research findings.

3.4 Data Analysis

The researchers systematically conducted screening procedures in SPSS to address potential issues related to missing values, straightlining, and outliers. Missing values were assessed using the count blank method to ensure data completeness. To detect straightlining, the researchers adhered to the guideline of ensuring a non-zero standard deviation value, as outlined by Hair et al. (2017). Additionally, an outlier analysis was performed using the Mahalanobis Distance test. Upon completing these initial screening steps, the researchers proceeded with a descriptive analysis.

For further analysis, the researchers employed Partial Least Squares Structural Equation Modelling (PLS-SEM) using SmartPLS software. This approach was selected for its ability to maximise the explained variance of exogenous variables on endogenous variables, prioritising predictive accuracy (Hair et al., 2017; Rigdon et al., 2017). PLS-SEM was particularly suited for the study's exploratory theoretical framework, as it is effective for testing additive causal models supported by theory, as recommended by Chin (1998) and Haenlein and Kaplan (2004). In contrast, Covariance-Based SEM (CB-SEM) focuses on assessing the alignment between a theoretical model and observed data, making it more applicable to confirmatory models (Ramayah et al., 2018). Originally developed by Wold (1974), SEM was further refined by his disciples Lohmöller and Jöreskog (1989), with Lohmöller (2013) adapting it into PLS-SEM and developing it into CB-SEM.

Before hypothesis testing, the researchers verified internal consistency, convergent validity, and discriminant validity. Internal consistency was assessed to ensure the interrelation among items, following Sekaran and Bougie (2016). A composite reliability value of at least 0.7 was required for successful implementation (Hair et al., 2017; Sekaran & Bougie, 2016). Convergent validity was examined through factor loadings and average variance extracted (AVE), which indicate the degree to which indicators of a construct share a substantial proportion of variance (Hair et al., 2017). Convergent validity was deemed satisfactory when factor loadings were at least 0.5 and the AVE score exceeded 0.5 (Byrne, 2013). Discriminant validity was confirmed to ensure the distinctiveness of each latent variable, indicating construct uniqueness (Hair et al., 2017; Ramayah et al., 2018). This was assessed using the heterotrait-monotrait ratio of correlations (HTMT), applying a threshold value of 0.9 as proposed by Henseler et al. (2015) and Franke and Sarstedt (2019). Furthermore, a lateral collinearity assessment was performed using a cut-off value of 3.3 to mitigate the risk of misleading results, as recommended by Diamantopoulos and Sigauw (2006). Following these rigorous validation procedures, the researchers proceeded with hypothesis testing using structural modelling techniques. This comprehensive methodological approach ensured the robustness and reliability of the study's findings.

4. Results

4.1 The Measurement Model

The structural equation modeling (SEM) technique, which establishes the relationships between measurement items and their corresponding latent variables, requires the development of a robust measurement model. This study lays the theoretical groundwork and employs statistical analyses to

ensure the validity and reliability of the measurement model. The analysis was conducted using the R programming language, leveraging the *lavaan* package for SEM analysis (Rosseel, 2012).

4.2 Operationalisation of Latent Variable

The conceptual model comprises four latent variables, which are inherently complex and cannot be measured by a single observed variable. To capture each latent variable within the conceptual model, the researchers employed multiple observed items. The first five exogenous variables pertain to teachers' grading practices (TGP). Measurement scales for TGP were developed based on the study by Link (2018). Under the domain of teachers' grading practices, 28 items were distributed across five dimensions: IMPC (Importance), USE (Usefulness), STUE (Student Effort), STUAB (Student Ability), and TGH (Teachers' Grading Habits). These items were adopted from prior literature to ensure robustness and relevance. For the measurement of Classroom-Based Assessment (CBA) practices, the researchers referred to foundational work by Chappuis et al. (2012) and adopted items from Goggin (2018). This multi-dimensional approach facilitated a comprehensive evaluation of the latent constructs in the study.

4.3 Normality Check

The researchers conducted both univariate and multivariate normality tests to assess whether the measurement items adhered to normal distribution before proceeding with the confirmatory factor analysis (CFA) model. The estimation method employed in CFA (and SEM) is contingent on the normality of the data, making it essential to ensure that the dataset meets this requirement. The results of the Shapiro-Wilk test (all p-values < 0.05) for all measurement items rejected the null hypothesis of univariate normality. Similarly, the Mardia test (p-value < 0.05) rejected the null hypothesis of multivariate normality. Given these findings, the researchers opted to use the maximum likelihood robust (MLR) estimator, also known as the Satorra-Bentler rescaling approach, for estimating the measurement model (Rosseel, 2012). This approach accommodates non-normal data, ensuring the robustness and reliability of the analysis.

Table 2

Measurement items and their reliability

Constructs and items	Factor loadings
Importance (IMPC, alpha: .882, CR: .894)	
1. IMPC1	0.736
2. IMPC2	0.844
3. IMPC3	0.741
4. IMPC4	0.861
5. IMPC5	0.832
6. IMPC6	0.729
Student Ability (STUAB, alpha: .892, CR: .894)	
1. STUAB1	0.644
2. STUAB2	0.713
3. STUAB3	0.780
4. STUAB4	0.800
5. STUAB5	0.770
6. STUAB6	0.849
7. STUAB7	0.846

Constructs and items	Factor loadings
Student Effort (STUE, alpha: .793, CR: .920)	
1. STUE1	0.634
2. STUE2	0.799
3. STUE3	0.694
4. STUE4	0.82
5. STUE5	0.607
Teacher Grading Practice (TGH, alpha: .807, CR: .821)	
1. TGH1	0.790
2. TGH2	Dropped
3. TGH3	0.601
4. TGH4	0.812
5. TGH5	0.748
6. TGH6	0.769
Usefulness (USE, alpha: .906, CR: .916)	
1. USE1	0.753
2. USE10	0.606
3. USE2	0.718
4. USE3	0.673
5. USE4	0.813
6. USE5	0.763
7. USE6	0.803
8. USE7	0.673
9. USE8	0.771
10. USE9	0.773
Classroom-based Assessment (CBA, alpha: .938, CR: .950)	
1. CBA1	0.635
2. CBA2	0.638
3. CBA3	0.828
4. CBA4	0.816
5. CBA5	0.738
6. CBA6	0.738
7. CBA7	0.641
8. CBA8	0.730
9. CBA9	0.760
10. CBA10	0.782
11. CBA11	0.792
12. CBA12	0.791
13. CBA13	0.716
14. CBA14	0.792

4.4 Reliability and Validity

The high outer loading values of the constructs demonstrate indicator reliability, suggesting a strong shared variance among the indicators. According to Hair et al. (2019), outer loading values should ideally be $\geq .708$; however, values as low as .700 are considered acceptable (Hair et al., 2017). Notably, in social science research, particularly when dealing with newly developed scales, weaker outer loading values (e.g., below .700) are not uncommon (Hulland, 1999). It is recommended to carefully evaluate the impacts on Composite Reliability (CR) and Average Variance Extracted (AVE) before excluding indicators with weaker outer loadings (Hair et al., 2017). In this study, only one indicator (TGH2) with the lowest loading value (.588) was removed, with consideration given to the resulting changes in AVE and CR. If the AVE value is $\geq .500$ and the CR value $> .700$, the researchers retained indicators even when their loading values were below .700. The decision to retain constructs aligns with Pallant's (2013) assertion that achieving high Cronbach's alpha values is challenging when the number of items in a construct is fewer than ten, making values $> .05$ acceptable. Furthermore, high composite reliability, high AVE, and moderate-to-high Cronbach's alpha indicate a well-specified measurement model in partial least squares structural equation modeling (PLS-SEM) (Hair et al., 2018; Hair et al., 2013). In this context, the researchers opted not to delete additional items. Thus, the reliability of the measurement model was confirmed.

Table 3

Divergent validity analysis

	CBA	IMPC	STUAB	STUE	TGH	USE
CBA						
IMPC	0.372					
STUAB	0.287	0.522				
STUE	0.324	0.488	0.869			
TGH	0.372	0.524	0.910	0.875		
USE	0.335	0.772	0.748	0.761	0.773	
AVE	0.556	0.628	0.600	0.513	0.559	0.544

Divergent or discriminant validity (DV) is established when constructs that are theoretically unrelated are statistically uncorrelated. DV can be confirmed by comparing the average variance extracted (AVE) with the squared correlations of all latent variables in a matrix, as demonstrated in Table 4. Following the guidelines of Hair et al. (2010), discriminant validity is indicated when the squared correlations below the diagonal are less than the AVE of each latent variable. Table 4 confirms that all constructs meet these criteria, establishing the discriminant validity of the latent variables. Furthermore, additional evidence for the accuracy and goodness-of-fit of the measurement model is provided by the Root Mean Square Error Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR), both of which are below the cut-off value of 0.08, as well as the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI), which exceed the recommended threshold of 0.90 (Hair et al., 2010). These results are presented in Table 2 and affirm the robustness of the measurement model. Following the establishment of the measurement model, the descriptive statistics are presented in Table 5.

Table 4

Descriptive statistics and association among latent variables

Latent Variables	Mean	SD	IMPC	USE	STUE	STUAB	TGH	CBA
IMPC	4.0247	0.49964	1.00					
USE	3.9917	0.53705	0.54	1.00				
STUE	3.7775	0.63130	0.60	0.61	1.00			
STUAB	3.8673	0.60900	0.64	0.62	0.67	1.00		
TGH	3.7288	0.59832	0.60	0.56	0.67	0.69	1.00	
CBA	3.8094	0.58957	0.64	0.54	0.66	0.67	0.71	1.00

It should be noted that the mean and standard deviation (SD) values are determined by taking the arithmetic mean of the item scores that measure the respective latent variables. Based on factor values that were retrieved by confirmatory factor analysis (CFA), the correlation matrix shows a correlation between the latent variables.

4.5 Assess the Structural Model for Collinearity Issue

The optimal VIF values, according to contemporary theories (Hair et al., 2019), should be lower than three (<3). For the PLS process, this study used Hair et al.'s (2019) threshold of VIF values of less than three to prevent collinearity and common method bias. Knock (2015, p. 7) posits that a VIF exceeding 3.3 signifies excessive collinearity, potentially indicating the presence of common method bias within the model. Consequently, if all VIFs derived from a thorough collinearity assessment are at or below 3.3, the model can be deemed devoid of common method bias. The SmartPLS specifically generated the order of predictors to evaluate the collinearity. IMPC, STUAB, STUE, TGH, and USE are predictors of CBA. The researchers assessed all VIF values; however, two indicators, STUAB (3.145) and USE (3.300), had values greater than 3. While the rest were assessed with the VIF values (<3): IMPC (2.032), STUE (2.908), and TGH (2.892). After considering the VIF values below, the researcher considered the work from Ringle et al. (2015) that concur collinearity issues are performed when the VIF value is equal to or greater than five. Therefore, VIF values >3.3 and <5 was accepted in this study. According to the results of all VIF values (Table 3), it can be noted that collinearity is not an issue in this study.

Table 5

VIF Values

	IMPC	STUAB	STUE	TGH	USE	CBA
CBA						
IMPC						2.032
STUAB						3.145
STUE						2.908
TGH						2.892
USE						3.300

Figure 1

The structural equation model that was estimated

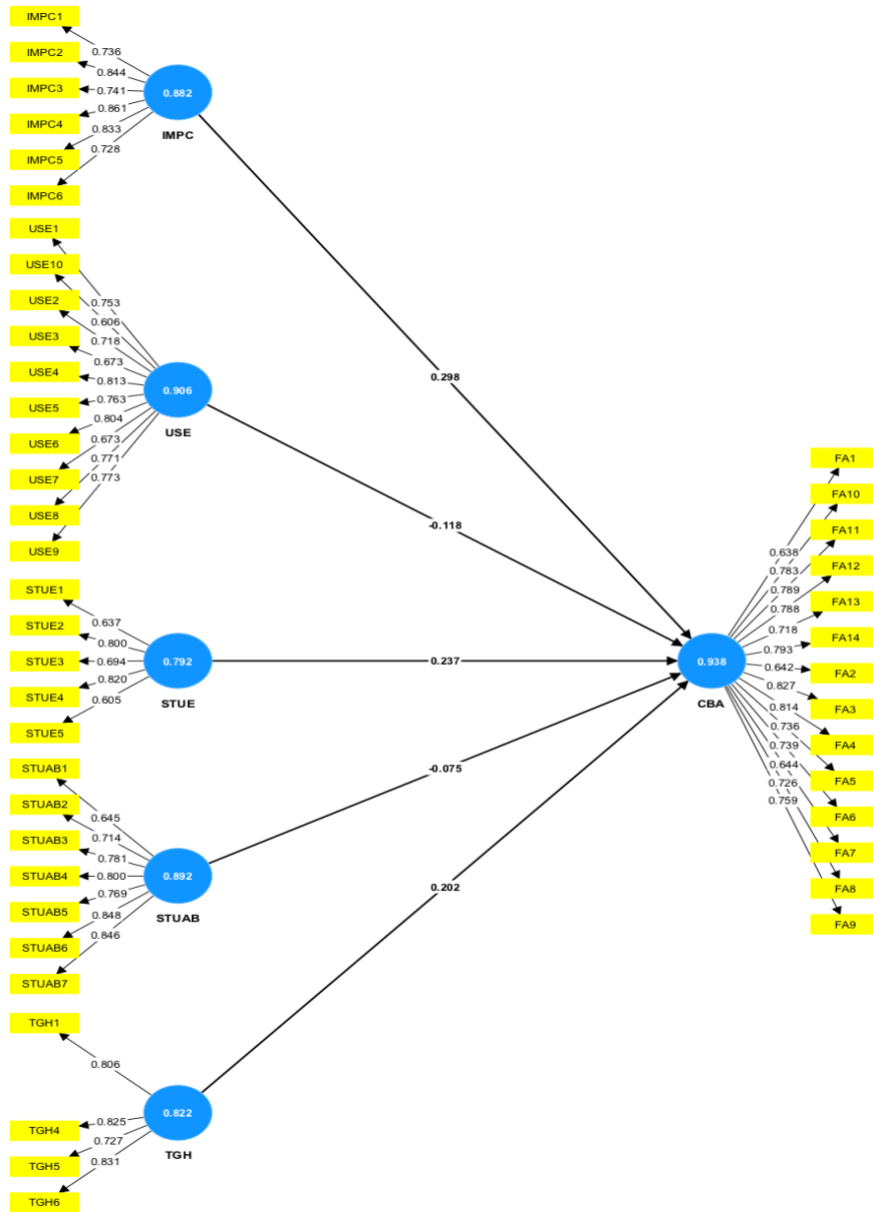


Table 6

Summary of hypothesis testing

Hypothesis	Standardized Coefficient	Remark
H1: IMPC -> CBA	.000	supported
H2: STUAB -> CBA	.173	Not supported
H3: STUE -> CBA	.004	supported
H4: TGH -> CBA	.002	supported
H5: USE -> CBA	.126	Not supported

Standard parenthesis error. ***p < 0.001, **p < 0.01, *p < 0.05, +p < 0.10.

4.6 Common Method Bias

Common method bias refers to measurement inaccuracies resulting from methodological issues in a study. An example of this bias is when the same measurement scale, such as a 5-point Likert scale, is used for all survey questions. In the present study, the researchers address common method bias by employing Harman's single-factor test, a widely used approach outlined by Podsakoff, et al. (2003). This test involves conducting an unrotated exploratory factor analysis on the 46 items loading on a single latent factor. The results indicate that only 39% of the average variance is explained by this single factor, significantly below the recommended cutoff point of 50%. Thus, based on these findings, common method bias is not a concern in this study.

4.7 Assess the Level of R²

The percentage of variance in the dependent variable explained by the model's independent variables is measured by the coefficient of determination (R²). An R² score of 0.8 or above is regarded as an adequate fit of the model to the data in Partial Least Squares Structural Equation Modelling (PLS-SEM), but an R² value of less than 0.5 indicates a poor match. The coefficient determination (R²) computation is shown in Table 7. The Classroom-Based Assessment (CBA) was found to have an R² coefficient of determination of 0.212, which indicates a reasonable level of explained variance of the dependent variable by the model. This shows a decent fit of the model to the data and a moderate level of prediction accuracy.

Table 7. Results of R² of the integrated model

	R ²	Consideration
CBA	.212	Moderate

4.8 Assessing the Predictive Relevance, Q²

The researchers evaluated the predictive significance of the model by using Blindfolded Techniques in SmartPLS 4 (Hair et al., 2019) to determine the Q² value. The Q² value assesses the model's ability to predict the dependent variable based on a latent variable. A high Q² value indicates robust predictive capacity for the hidden variable in forecasting the dependent variable, while a low Q² value suggests limited predictive ability. Table 8 presents the results of the predictive relevance evaluation, and it shows that the Q² value for CBA is above 0 (.175), confirming the predictive validity of the model's endogenous construct.

Table 8

Results of Q² of the integrated model

	Q ²	Predictive relevance
CBA	.175	Moderate

5. Discussion

Grading practices and classroom-based assessment practices are inextricably linked because grades are frequently based on assessments given in the classroom. In determining CBA practices as the main objective of this part, this study reports that only three hypothesised variables (IMPC, STUE, and TGH) are statistically significant in predicting CBA practices. The other two factors (STUAB and USE) are not statistically significant in predicting CBA practices. In detail, Table 6 above presents the three hypotheses, their coefficients and t-values, and the decision over the matters for Grading Practices towards CBA practices (H1–H5).

In predicting preservice teachers' CBA practices during teaching and learning sessions, importance (IMPC), which refers to a specific aspect or component that preservice teachers consider significant in grading practices, was reported to be substantial in predicting classroom-based assessment practices. Specifically, it indicates the factors or elements preservice teachers prioritise or attach particular importance to when evaluating students' performance and assigning grades. A similar result was found by Guskey and Link (2019) and Karlen et al. (2024), where they observed an association between Importance and assessment practice. The possible explanation for this finding is that preservice teachers may view specific criteria or skills as crucial for achieving the learning objectives of their courses. As a result, they prioritise assessing and grading students' performance based on these critical dimensions that directly relate to the intended outcomes (Mat Yusoff et al., 2023).

Not only that, Student Effort (STUE), which refers to students' level of engagement, motivation, and hard work in their learning, was reported to be the most decisive factor in the Grading Practices variable, followed by Teachers' Grading Habits (TGH). This finding could be interpreted based on the assumption that when students participate in meaningful and relevant assessments, they are much more likely to put in the effort necessary to learn the material (Mohamad Marzaini et al., 2024). This is because preservice teachers believe that their students understand the importance and relevance of the assessments to their learning and long-term goals. Furthermore, well-designed assessments aligned with learning goals offer students reasonable expectations and a sense of meaning (Andrade & Brookhart, 2020). The findings from this study align with prior investigations that indicate a significant correlation between STUE and CBA practices (Lee & Jeon, 2024; Andrade & Brookhart, 2020; Andersson & Palm, 2017).

Additionally, the outcomes of this study disclosed that Teachers' Grading Habits (TGH)—the specific strategies and methods preservice teachers employ for evaluating and assigning grades to students' work—influence CBA practices. Similar findings were presented by Hung and Wu (2024), Guskey and Link (2019), Knight and Cooper (2019), and Chen and Bonner (2017). According to Swan Sein et al. (2021), preservice teachers with well-established grading habits are much more likely to use assessments associated with learning objectives and provide precise and equitable evaluations of student learning. Thus, they are also more inclined to use clear and uniform criteria for assessing and providing feedback on student work, which can assist students in comprehending their weaknesses and strengths and enhance their learning. Besides, Mat Yusoff et al., (2025) and Chamberlin et al. (2018) illustrated that preservice teachers' grading habits can influence student engagement and motivation in the classroom. Therefore, preservice teachers who use meaningful and relevant assessments and provide regular and meaningful feedback are prone to captivate students and encourage them to put in more effort (Mat Yusoff et al., 2024; Swan Sein et al., 2021). Furthermore, these preservice teachers may believe that their grading habits impact how they provide feedback to students and communicate the value of student effort. For example, if a preservice teacher has a habit of providing lots of feedback and encouragement, this can result in more student motivation and engagement.

On the contrary, Student Ability (STUAB) and Usefulness (USE) were found to be not significant in predicting CBA practices during teaching and learning sessions. While USE is an important consideration for preservice teachers when designing assessments, they might have little influence on the types of assessments employed or the methods employed to assess student learning (Mat Yusoff et al., 2023; Schildkamp et al., 2017). A preservice teacher may believe it is vital to assess a specific skill, but they may not have the resources or time to do so during a particular teaching and learning session. Furthermore, while a preservice teacher may find a specific assessment beneficial for student learning, it might not suit particular learning objectives or student needs (Radović et al., 2024; Guskey & Link, 2019).

Second, STUAB is also found to be insignificant in predicting CBA practices. This is because student ability is not the only factor influencing student performance; other aspects, including student effort, learning strategies, and motivation, could also play a role (Mat Yusoff et al., 2023; Sharma & Sharma, 2018). This could also be assumed by preservice teachers, who may believe that a motivated student will be more engaged in the learning process and more likely to put in the effort required to succeed. Even if a student's ability is limited, if they are motivated to learn, they are seen to succeed more than a student with more remarkable ability but low motivation. Furthermore, while student ability is unquestionably essential, it is not a static trait and can be modified and developed over time (Abdul Razak et al., 2023; Guskey & Link, 2019). Preservice teachers can use different assessments and

strategies to enable students to develop their abilities, including providing opportunities for practice, feedback, and revision, which can help students enhance their skills over time (Mat Yusoff et al., 2024; Knight & Cooper, 2019).

The Grading Practices factor is not significant in predicting classroom-based assessment practice in the Malaysian context because Malaysia's education system adheres to a centralised curriculum and assessment policy, which governs the types of assessments used, the criteria for assessing student work, and the techniques employed to provide pupils with feedback (Abdul Razak et al., 2023). Preservice teachers' ability to use their grading practices is limited, and it may be challenging to align their grading practices with their CBA practices. Also, Malaysia's education system places a high value on high-stakes tests like the UPSR, PT3, and SPM, which are used to ascertain students' qualifications for further education (Mohamad Marzaini et al., 2024). As a result, preservice teachers have shifted their focus to preparing students for these tests instead of employing various assessment techniques that correspond to learning goals in addition to providing accurate and fair evaluations of student learning.

6. Conclusion

This study explored the relationship between preservice teachers' grading practices (TGP) and classroom-based assessment (CBA) practices among preservice teachers in Selangor, Malaysia. The findings provided valuable insights into the factors that significantly influence the implementation of CBA and highlighted areas requiring improvement in grading practices. The results of the study indicated that among the five dimensions of TGP, three were found to be significant predictors of CBA practices: Importance (IMPC), Student Effort (STUE), and Teachers' Grading Habits (TGH). These dimensions highlighted the key factors preservice teachers prioritise when evaluating students' performance and assigning grades. Preservice teachers placed particular importance on specific criteria or skills that directly relate to the intended learning outcomes, emphasising the need for assessments aligned with these critical dimensions.

Student Effort (STUE) emerged as a crucial factor influencing CBA practices. Preservice teachers recognised that students' engagement, motivation, and hard work played a significant role in their learning. When assessments were meaningful and relevant, students were likelier to put in the effort necessary to succeed. Moreover, Teachers' Grading Habits (TGH) impacted CBA practices, indicating that well-established grading habits, including clear and uniform criteria for assessment and providing regular and meaningful feedback, contributed to enhanced student engagement and motivation. On the other hand, the dimensions of Student Ability (STUAB) and Usefulness (USE) were not significant predictors of CBA practices. This suggests that while these factors are important considerations, they may have less influence on the types of assessments employed or the methods used to assess student learning. Additionally, Malaysia's centralised curriculum and assessment policy may limit preservice teachers' flexibility in implementing their grading practices, particularly concerning CBA.

The findings of this study contribute to the existing knowledge on grading practices and CBA, specifically within the context of Malaysian preservice teachers. By identifying the significant factors influencing CBA practices, this research provides valuable insights for educators and policymakers in refining assessment procedures, enhancing fairness, and fostering an optimal learning environment. It underscores the importance of aligning assessments with intended learning outcomes, emphasising student effort and engagement, and cultivating effective grading habits.

To bridge the gap between recommended grading practices and actual implementation, targeted interventions and professional development programs can be designed to support preservice teachers in enhancing their assessment strategies and aligning them with CBA goals. These initiatives can focus on developing preservice teachers' pedagogical content knowledge, promoting unbiased decision-making, and improving lesson planning to ensure accurate reporting of student performance. Overall, this study contributes to the knowledge of grading practices and CBA, providing a basis for future research and improvement in the Malaysian educational landscape. By addressing the validity issues in grading practices and promoting the effective implementation of CBA, educators can enhance student learning experiences, engagement, and motivation, ultimately leading to improved academic achievements and educational opportunities.

7. Co-Author Contribution

All authors have contributed significantly to this study. Shahazwan Mat Yusoff conceptualized the research framework, supervised data collection, and led the manuscript preparation. Anwar Farhan Mohamad Marzaini and Siti Nadya Zynuddin contributed to the literature review, methodological design, and data analysis. Ahmad Muhaimin Mat Jusoh was responsible for statistical analysis and interpretation of results. Shahazwan Mat Yusoff and Noorhayati Zakaria provided critical revisions and theoretical insights and ensured the alignment of findings with the broader academic discourse. All authors participated in manuscript drafting, reviewed the final version, and approved it for publication.

8. Acknowledgements

The authors would like to express their deepest gratitude to all individuals and organizations who contributed to the successful completion of this study. Their cooperation was instrumental in ensuring the study's success. Additionally, appreciation is extended to the participants who generously shared their experiences and colleagues and mentors who provided guidance and encouragement throughout the research process.

9. References

- Abd Mutalib, S., & Ahmad, J. (2012). Kepelbagaian Aras Soalan Lisan dalam Pentaksiran Formatif bagi Subjek Geografi. *Sains Humanika*, 59(1), 103-108.
- Abdul Razak, R., Mat Yusoff, S., Hai Leng, C., & Mohamadd Marzaini, A. F. (2023). Evaluating teachers' pedagogical content knowledge in implementing classroom-based assessment: A case study among esl secondary school teachers in Selangor, Malaysia. *Plos one*, 18(12), e0293325.
- Ahmad, A., & Mahamod, Z. (2016). Tahap Kemahiran Guru Bahasa Melayu Sekolah Menengah Dalam Melaksanakan Pentaksiran Berasaskan Sekolah Berdasarkan Jantina, Opsyen Dan Tempat Mengajar (Level of Malay Language Secondary School Teacher in the Implementation School Based Assessment Towards Gender, Option and Teaching Location). *Jurnal Pendidikan Bahasa Melayu*, 5(1), 18-29.
- Alam, A., & Mohanty, A. (2023). Cultural beliefs and equity in educational institutions: exploring the social and philosophical notions of ability groupings in teaching and learning of mathematics. *International Journal of Adolescence and Youth*, 28(1), 2270662.
- Ambotang, A. S., & Gobalakrishnan, S. (2017). Tahap Integriti dan Kesiediaan Guru dengan Pelaksanaan Pentaksiran Berasaskan Sekolah. *International Journal of Education, Psychology Counseling*, 2(4), 1-22.
- Andersson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and instruction*, 49, 92-102.
- Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. *Assessment in education: principles, policy & practice*, 27(4), 350-372.
- Ansori, M., Nurkamto, J., & Suparno. (2019). Teacher's beliefs and practices in the integration of higher order thinking skills in teaching reading. *ELS Journal on Interdisciplinary Studies in Humanities*, 2(4), 541-555. DOI: [10.34050/els-jish.v2i4.8164](https://doi.org/10.34050/els-jish.v2i4.8164)
- Bidin, N., & Mahamod, Z. (2016). Kesiediaan Murid Sekolah Menengah Terhadap Pelaksanaan Pentaksiran Berasaskan Sekolah Dalam Mata Pelajaran Bahasa Melayu (Readiness Secondary School Students on the Implementation of School Based Assessment in Malay Language Subject). *Jurnal Pendidikan Bahasa Melayu*, 6(1), 64-76.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7-74.
- Brewer, T. J., & deMarrais, K. (2015). Teach for America counter-narratives: Alumni speak up and speak out. New York, NY: Peter Lang.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35-36.

- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123-142.
- Brookhart, S. M., & Nitko, A. J. (2019). *Educational assessment of students* (8th ed.). Boston, MA: Pearson.
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of educational research, 86*(4), 803-848.
- Bürgermeister, A., Glogger-Frey, I., & Saalbach, H. (2021). Supporting peer feedback on learning strategies: Effects on self-efficacy and feedback quality. *Psychology Learning & Teaching, 20*(3), 383-404.
- Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. routledge.
- Chamberlin, K., Yasué, M., & Chiang, I. C. A. (2023). The impact of grades on student motivation. *Active Learning in Higher Education, 24*(2), 109-124.
- Chappuis, J., Stiggins, R. J., Chappuis, S., & Arter, J. (2020). *Classroom assessment for student learning: Doing it right--using it well* (p. 432). New York: Pearson.
- Che Aleha Ladin & Md Nasir Ibrahim.(2015). Pentaksiran Berasaskan Sekolah Pendidikan Seni Visual Sekolah Menengah di Malaysia.Imajinasi: Jurnal Seni,9(1),63-70.
- Chen, P. P., & Bonner, S. M. (2017). Teachers' beliefs about grading practices and a constructivist approach to teaching. *Educational Assessment, 22*(1), 18-34.
- Cheng, L., DeLuca, C., Braund, H., & Yan, W. (2018). Grading policies in Canada and China: A comparative study. *Comparative and International Education, 47*(1), 1-25.
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research, 295*(2), 295-336.
- Cizek, G. J., Andrade, H. L., & Bennett, R. E. (2019). Formative assessment. *Handbook of formative assessment in the disciplines, 3-19*.
- Collins, R. (2014). Skills for the 21st Century: teaching higher-order thinking. *Curriculum & Leadership Journal, 12*(14), 1-8.
- Coombs, A., Ge, J., & DeLuca, C. (2021). From sea to sea: The Canadian landscape of assessment education. *Educational Research, 63*(1), 9-25.
- Creswell, J. W. (2014). *A concise introduction to mixed methods research*. SAGE publications.
- Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British journal of management, 17*(4), 263-282.
- Dyrness, R., & Dyrness, A. (2008). Making the grade in middle school. *Kappa Delta Pi Record, 44*(3), 114-118.
- Fadila, N., Mahamod, Z., & Mohammad, W. M. R. W. (2016). Kompetensi, Sikap Dan Penggunaan Aplikasi Web 2.0 Sebagai Bahan Bantu Mengajar Dalam Kalangan Guru Bahasa Melayu Sekolah Rendah (Competencies, Attitude and Application of Web 2.0 as a Teaching Aids among Malay Language Primary School Teachers). *Jurnal Pendidikan Bahasa Melayu, 6*(2), 52-58.
- Faizah, A. M. (2011). School-based assessment in Malaysian schools: The concerns of the English teachers. *US-China Education Review B3*(2011), 393-402.
- Feldman, J. (2018). *Grading for equity: What it is, why it matters, and how it can transform schools and classrooms*. Corwin Press.
- Franke, G., & Sarstedt, M. (2019). Heuristics versus statistics in discriminant validity testing: a comparison of four procedures. *Internet Research, 29*(3), 430-447.
- Gaertner, M. N. (2022). *Norm-Referenced Assessment*. Routledge.
- Gentrup, S., Lorenz, G., Kristen, C., & Kogan, I. (2020). Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement. *Learning and Instruction, 66*, 101296.
- Getenet, S., Haeusler, C., Redmond, P., Cante, R., & Crouch, V. (2024). First-year Preservice Teachers' Understanding of Digital Technologies and Their Digital Literacy, Efficacy, Attitude, and Online Learning Engagement: Implication for Course Design. *Technology, Knowledge and Learning, 29*, 1359-1383.

- Goggin, S. E. (2018) *A qualitative study of the implementation formative assessment strategies in the classroom* (Doctoral dissertation). <https://www.proquest.com/docview/2551278456?pq-origsite=gscholar&fromopenview=true>.
- Green, C. L., Walker, J. M., Hoover-Dempsey, K. V., & Sandler, H. M. (2007). Parents' motivations for involvement in children's education: An empirical test of a theoretical model of parental involvement. *Journal of educational psychology*, 99(3), 532-544.
- Guskey, T. R. (2011). Five obstacles to grading reform. *Educational Leadership*, 69(3), 16-21.
- Guskey, T. R., & Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice*, 26(3), 303-320.
- Guskey, T. R., & Link, L. J. (2019). Exploring the factors teachers consider in determining students' grades. *Assessment in Education: Principles, Policy & Practice*, 26(3), 303-320.
- Haenlein, M., & Kaplan, A. M. (2004). A beginner's guide to partial least squares analysis. *Understanding statistics*, 3(4), 283-297.
- Hair Jr, J. F., Matthews, L. M., Matthews, R. L., & Sarstedt, M. (2017). PLS-SEM or CB-SEM: updated guidelines on which method to use. *International Journal of Multivariate Data Analysis*, 1(2), 107-123.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. *Long range planning*, 46(1-2), 1-12.
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European business review*, 31(1), 2-24.
- Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010) *Multivariate Data Analysis*, 7th ed., Prentice Hall, Upper Saddle River, NJ.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, 43, 115-135.
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., ... & Vedel, I. (2019). Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. *Journal of clinical epidemiology*, 111, 49-59.
- Horsburgh, J., & Ippolito, K. (2018). A skill to be worked at: using social learning theory to explore the process of learning from role models in clinical settings. *BMC medical education*, 18(1), 1-8.
- Huberman, A. M., & Miles, M. B. (1994). Data management and analysis methods.
- Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic management journal*, 20(2), 195-204.
- Hung, C. S., & Wu, H. K. (2024). High school science teachers' assessment literacy for inquiry-based science instruction. *International Journal of Science Education*, 46(7), 621-642.
- Isa, N. M., & Naim, H. A. (2013). Kajian keperluan pembinaan modul pentaksiran berasaskan sekolah bagi tajuk dinamik dalam mata pelajaran sains tingkatan 2. In *2nd International Seminar on Quality and Affordable Education*.
- Kaplan, R., James, J., Figueroa, D. T., Rawkins, C., & Dumont, C. (2020). Education policy outlook Estonia-June 2020.
- Karlen, Y., Bäuerlein, K., & Brunner, S. (2024). Teachers' assessment of self-regulated learning: Linking professional competences, assessment practices, and judgment accuracy. *Social Psychology of Education*, 27(2), 461-491.
- Knight, M., & Cooper, R. (2019). Taking on a new grading system: The interconnected effects of standards-based grading on teaching, learning, assessment, and student behavior. *NASSP Bulletin*, 103(1), 65-92.
- Knight, M., & Cooper, R. (2019). Taking on a new grading system: The interconnected effects of standards-based grading on teaching, learning, assessment, and student behavior. *NASSP Bulletin*, 103(1), 65-92.
- Knock, N. (2015). Common method bias in PLS-SEM. *International Journal of E-Collaboration*, 11(4), 1-10.
- Kubiszyn, T., & Borich, G. D. (2016). *Educational testing and measurement*. John Wiley & Sons.
- Kulasegaram, K., & Rangachari, P. K. (2018). Beyond "formative": assessments to enrich student

- learning. *Advances in physiology education*, 42(1), 5-14.
- Kunnath, J. P. (2017). Teacher grading decisions: Influences, rationale, and practices. *American Secondary Education*, 45(3), 68-88.
- Lee, S., & Jeon, J. (2024). Teacher agency and ICT affordances in classroom-based language assessment: The return to face-to-face classes after online teaching. *System*, 121, 103218.
- Lewkowicz, J., & Leung, C. (2021). Classroom-based assessment. *Language Teaching*, 54(1), 47-57.
- Lijie, H., Mat Yusoff, S., & Mohamad Marzaini, A. F. (2024). Influence of AI-driven educational tools on critical thinking dispositions among university students in Malaysia: a study of key factors and correlations. *Education and Information Technologies*, 1-25.
- Link, L. (2018). Teachers' Perceptions of Grading Practices: How Pre-Service Training Makes a Difference. *Journal of Research in Education*, 28(1), 62-91.
- Link, L. (2018). Teachers' Perceptions of Grading Practices: How Pre-Service Training Makes a Difference. *Journal of Research in Education*, 28(1), 62-91.
- Linton, J. N. (2017). Institutional factors for supporting electronic learning communities. *Online Learning*, 21(1), 238-256.
- Liu, X. (2008). Assessing measurement invariance of the teachers' perceptions of grading practices scale across cultures. *NERA Conference Proceedings 2008*. 3.
https://digitalcommons.lib.uconn.edu/nera_2008/3
- Lohmöller, J. B. (2013). *Latent variable path modeling with partial least squares*. Springer Science & Business Media. *springer*
- Lohmöller, JB. (1989). Predictive vs. Structural Modeling: PLS vs. ML. In: Latent Variable Path Modeling with Partial Least Squares. Physica, Heidelberg.
https://doi.org/10.1007/978-3-642-52512-4_5
- Lokman, H. F., Abdullah, N., Mustafa, M. C., Hanapi, M. H. M., Sembak, S., & Baskaran, V. L. (2024). The Impact of Classroom-Based Assessment (CBA) Implementation on the Performance of B40 Pre-School and Primary School Students. *International Journal of Learning, Teaching and Educational Research*, 23(7), 435-460.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing research*, 35(6), 382-386.
- Mahalingappa, L. (2024). Development of linguistic critical consciousness of multilingual preservice teachers of color. *International Multilingual Research Journal*, 18(3), 197-207.
- Mansor, A. N., Vikaraman, S. S., Medina, N. I., & Alias, B. S. (2019). Managing school-based assessment: Challenges and solutions for educational practice. *International Journal of Innovation, Creativity and Change*, 7(7), 63-84.
- Mat Yusoff, S., Marzaini, A. F. M., Zakaria, N., & Lijie, H. (2025). Exploring the Implementation of Classroom-Based Assessment in Malaysian Secondary Schools: Alignment With Policy and Teacher Practices. *MOJEM: Malaysian Online Journal of Educational Management*, 13(1), 81-101.
- Mat Yusoff, S., Marzaini, A. F. M., Hassan, M. H., & Zakaria, N. (2023). Investigating the Roles of Pedagogical Content Knowledge in Music Education:: A Systematic Literature Review. *Malaysian Journal of Music*, 12(2), 95-114.
- Mat Yusoff, S., Leng, C. H., Razak, R. A., Basal, M. H., Marzaini, M., & Farhan, A. (2024). Conceptions of Assessment: Perceptions of Physical and Health Education Teachers in Selangor, Malaysia. *Pertanika Journal of Social Sciences & Humanities*, 32(1), 189-216..
- Mat Yusoff, S., Lijie, H., Marzaini, A. F. M., & Basal, M. H. (2024). An Investigation of the Theory of Planned Behavior in Predicting Malaysian Secondary School Teachers' Use of ICT during Teaching and Learning Sessions. *Journal of Nusantara Studies (JONUS)*, 9(1), 97-120.
- Mat Yusoff, S., Razak, R. A., Leng, C. H., & Marzaini, A. F. (2023). Determining The Conceptions of Assessment among Secondary School In-Service Science Teachers. *MOJES: Malaysian Online Journal of Educational Sciences*, 11(3), 57-70.
- Matthewman, G., Lee, A., Kaur, J. G., & Daley, A. J. (2018). Physical activity for primary dysmenorrhea: a systematic review and meta-analysis of randomized controlled trials. *American journal of obstetrics and gynecology*, 219(3), 255.e1-255.e20.
- Mazlini Adnan & Noorfazelawati Abd Kadir. (2014). The Practice Of School Based Assessment (SBA)

- Among Secondary School Mathematics Teachers. *Jurnal Pendidikan Sains dan Matematik Malaysia*, 4(1), 59-69.
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational horizons*, 87(1), 40-49.
- McMillan, J. H., & Hellsten, L. (2010). *Classroom assessment: Principles and practice for effective standards-based instruction*. Pearson Education Canada.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20-32.
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The journal of educational research*, 95(4), 203-213.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational researcher*, 18(2), 5-11.
- Mohamad Marzaini, A. F., Yusoff, S. M., Zakaria, N., Basal, M. H., Sharil, W. N. E. H., & Supramaniam, K. (2024). A Discovery On Malaysian ESL Teachers' adherence to Policy Mandates in Classroom Based Assessment Practices. *Malaysian Journal of Learning and Instruction*, 21(1), 249-284.
- Moktar, S. N. A., Hanapi, Z., Kiong, T. T., Mohamed, S., & Rus, R. C. (2018). Kesiediaan, Penerimaan Dan Pengoperasian Guru Reka Bentuk dan Teknologi Terhadap Amalan dalam Pentaksiran Berasaskan Sekolah. *Sains Humanika*, 10(3-3), 89-94.
- Narinasamy, I., & Nordin, N. A. (2018). Implementing Classroom Assessment in Malaysia: An Investigation. *J. Kurikulum*, 3, 55-63.
- O'Connor, K. (2007). A repair kit for grading. *Portland, OR: Educational Testing Service*.
- Olsen, B., & Buchanan, R. (2019). An investigation of teachers encouraged to reform grading practices in secondary schools. *American Educational Research Journal*, 56(5), 2004-2039.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational research review*, 9, 129-144.
- Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018). Implementation of formative assessment—effects of quality of programme delivery on students' mathematics achievement and interest. *Assessment in Education: Principles, Policy & Practice*, 25(2), 160-182.
- Pirzad, F., & Abadikhah, S. (2022). An Evaluation of Iranian Tenth-grade English Textbook: With a Focus on Language Skills, Activities and Teachers' Perceptions. *Two Quarterly Journal of English Language Teaching and Learning University of Tabriz*, 14(29), 171-197.
- Pitt, E., & Carless, D. (2022). Signature feedback practices in the creative arts: integrating feedback within the curriculum. *Assessment & Evaluation in Higher Education*, 47(6), 817-829.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. <https://doi.org/czw>
- Ramayah, T., Cheah, J., Chuah, F., Ting, H., & Memon, M. A. (2018). Partial least squares structural equation modeling (PLS-SEM) using SmartPLS 3.0: An updated guide and practical guide to statistical analysis (2nd ed.). Kuala Lumpur, Malaysia: Pearson.
- Randall, J., & Engelhard, G. (2009). Differences between teachers' grading practices in elementary and middle schools. *The Journal of Educational Research*, 102(3), 175-186.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372-1380.
- Redding, C., & Smith, T. M. (2016). Easy in, easy out: Are alternatively certified teachers turning over at increased rates?. *American Educational Research Journal*, 53(4), 1086-1125.
- Reeves, D. (2011). *Elements of grading: A guide to effective practice*. Solution Tree Press.
- Reynolds, M. C. (1989). *Knowledge base for the beginning teacher*. Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523.
- Rigdon, E. E., Sarstedt, M., & Ringle, C. M. (2017). On comparing results from CB-SEM and PLS-SEM: Five perspectives and five recommendations. *Marketing: ZFP—Journal of Research and Management*, 39(3), 4-16.
- Ringle, C., Da Silva, D., & Bido, D. (2015). Structural equation modeling with the SmartPLS. *Bido, D., da Silva, D., & Ringle, C. (2014). Structural Equation Modeling with the Smartpls. Brazilian Journal Of Marketing*, 13(2), 56-73.

- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of statistical software*, 48, 1-36.
- Russell, J. A., & Austin, J. R. (2010). Assessment practices of secondary music teachers. *Journal of research in music education*, 58(1), 37-54.
- Schildkamp, K., Poortman, C., Luyten, H., & Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. *School effectiveness and school improvement*, 28(2), 242-258.
- Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*. John Wiley & sons.
- Sharma, D., & Sharma, S. (2018). Relationship between motivation and academic achievement. *International Journal of Advances in Scientific Research*, 4(1), 1-5.
- Shepard, L. A. (2017). Formative assessment: Caveat emptor. In *The future of assessment* (pp. 279-303). Routledge.
- Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 183-200.
- Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational measurement: issues and practice*, 37(1), 21-34.
- Singleton, R. A., & Straits, B. C. (2012). Survey interviewing. *The SAGE handbook of interview research: The complexity of the craft*, 77-98.
- Smith, M. A. (2014). *Instructional and grading practices that change grading fidelity: An action research study* (Doctoral dissertation, Capella University).
- Stiggins, R. J., Schafer, W. D., & Hills, T. D. (2005). Assessment literacy, Phi Del Kappan 83(10), 758-765.
- Swan Sein, A., Rashid, H., Meka, J., Amiel, J., & Pluta, W. (2021). Twelve tips for embedding assessment for and as learning practices in a programmatic assessment system. *Medical Teacher*, 43(3), 300-306.
- Tai, J., Ajjawi, R., Bearman, M., Boud, D., Dawson, P., & Jorre de St Jorre, T. (2023). Assessment for inclusion: Rethinking contemporary strategies in assessment design. *Higher Education Research & Development*, 42(2), 483-497.
- Tierney, R. D., Simon, M., & Charland, J. (2011, July). Being fair: Teachers' interpretations of principles for standards-based grading. In *The educational forum* (Vol. 75, No. 3, pp. 210-227). Taylor & Francis Group.
- Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review*, 32, 100374.
- Varatharaj, R., Abdullah, A. G. K., & Ismail, A. (2015). The effect of teacher autonomy on assessment practices among Malaysian cluster school teachers. *International Journal of Asian Social Science*, 5(1), 31-36.
- Vykydal, D., Foltá, M., & Nenadál, J. (2020). A study of quality assessment in higher education within the context of sustainable development: A case study from Czech Republic. *Sustainability*, 12(11), 4769.
- Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39(1), 1-37.
- Wold, H. (1974). Causal flows with latent variables: partings of the ways in the light of NIPALS modelling. *European economic review*, 5(1), 67-86.
- Wormeli, R. (2006). Accountability: Teaching through assessment and feedback, not grading. *American secondary education*, 34(3), 14-27.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, policy & practice*, 17(1), 59-75.
- Yesbeck, D. M. (2011). *Grading practices: Teachers' considerations of academic and non-academic factors*. Virginia Commonwealth University ProQuest Dissertations & Theses, 2011. 3487509.
- Yuh, T. J., & Kenayathulla, H. B. (2020). Pentaksiran bilik darjah dan prestasi murid sekolah jenis

kebangsaan cina di Hulu Langat, Selangor. *JuPiDi: Jurnal Kepimpinan Pendidikan*, 7(3), 70-90.

Zapata-Rivera, D., Jackson, T., & Katz, I. (2015). Authoring conversation-based assessment scenarios. *Design recommendations for intelligent tutoring systems*, 3, 169-178.