

Institutionalizing classroom-based assessment: ESL teachers' grading practices in Malaysian secondary schools

Anwar Farhan Mohamad Marzaini¹, Shahazwan Mat Yusoff¹, Muhammad Usamah Mohd Ridzuan²,
Nur Fatin Shahmina Mohd Fauzey², Nur Husna Serip Mohamad², Noraziah Mohd Amin²,
Muhammad Aiman Abdul Halim², Syafiq Munir Ismail Munir³

¹Department of Curriculum and Instructional Technology, Faculty of Education, Universiti Malaya, Kuala Lumpur, Malaysia

²Academy of Language Studies, Universiti Teknologi MARA Cawangan Pulau Pinang, Pulau Pinang, Malaysia

³Department of Science and Technology Studies, Faculty of Science, Universiti Malaya, Kuala Lumpur, Malaysia.

Article Info

Article history:

Received Dec 1, 2024

Revised Feb 8, 2026

Accepted Mar 30, 2026

Keywords:

Classroom-based assessment

ESL teachers

Grading practices

Malaysian schools

Policy mandates

ABSTRACT

Effective assessment practices play a pivotal role in supporting student learning, particularly within the framework of classroom-based assessment (CBA). This study investigates how English as a second language (ESL) teachers in Malaysia grade students during CBA and their alignment with policy mandates. Employing a qualitative phenomenological design, data were gathered through semi-structured interviews and analysis of curriculum documents from the CBA Guidebook. Six Form 2 ESL teachers from three public secondary schools in Pulau Pinang participated in the study. Thematic analysis revealed commendable grading practices, with teachers employing various strategies. However, a misalignment was noted between these practices and policy directives, as teachers tended to adopt exam-oriented approaches. The study has implications for ESL education in Malaysia, highlighting the need for better alignment between policy and practice and providing practical recommendations to align grading methods with educational objectives. This study contributes to understanding how ESL teachers interpret and enact assessment reforms, offering practical insights for aligning grading with formative policy goals.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Anwar Farhan Mohamad Marzaini

Department of Curriculum and Instructional Technology, Faculty of Education, Universiti Malaya

50603, Kuala Lumpur, Malaysia.

Email: are_not46@yahoo.com

1. INTRODUCTION

The Malaysian education system has long been shaped by an exam-focused culture, emphasizing high-stakes assessment like the *Ujian Pencapaian Sekolah Rendah* (UPSR), *Penilaian Tingkatan 3* (PT3), and *Sijil Pelajaran Malaysia* (SPM) [1], [2]. This focus has been criticized for limiting meaningful learning, especially among students who struggle with exams [3]. In response, the ministry of education launched classroom-based assessment (CBA) reforms in 2021, aiming to integrate formative approaches into a holistic, less stressful evaluation system [4]. Teachers' responsibilities now span planning, implementation, assessment, evaluation, reporting, and follow-up actions, as dictated by policy guidelines [5]. This change significantly reconfigures the role of grading in Malaysian classrooms. Grading has shifted from a one-off exam score to an ongoing student development, teacher feedback, and performance evidence across multiple dimensions [6]. Internationally, the shift from high-stakes exam to formative assessment has faced challenges. In countries like Finland and Scotland, reforms emphasized teacher autonomy, yet policy clarity and consistency remained issues [7]. New Zealand required years of professional development to achieve grading consistency under CBA

[8]. In Malaysia, these challenges are intensified by a longstanding exam-oriented culture, where grading now involves balancing personal judgment, pedagogy, student evidence, and top-down policy demands [8]. In this context, effective grading under formative systems demands critical thinking and reflective decision-making skills that develop over time through training and experience [9]. Yet, Marnizam and Ali [2] noted that these decisions are often fraught with tensions between internal beliefs and external mandates. Although Malaysia's CBA initiative has been in place since 2019, many teachers report a lack of clarity and training on how to grade within this new assessment culture [10]. The short rollout period, minimal teacher consultation, and limited professional development have hindered smooth execution. Scholars suggest that sustainable educational reform typically requires 8 to 14 years to mature [11]. This raised questions about whether Malaysia's CBA reform has been adequately supported. While previous studies examined teachers' assessment literacy [2], readiness for reform [12], [13], and general assessment strategies [5], [14], limited attention has been paid to how grading practices align-or fail to align-with CBA policy expectations. Existing studies tend to describe systemic issues or conceptual understanding, but they rarely capture the fine-grained, day-to-day decision-making through which teachers interpret policy, judge student evidence, and negotiate tensions between their personal beliefs and official expectations. By foregrounding teacher sense-making and the practical enactment of grading in real classrooms, this study offers a more nuanced understanding of how CBA reforms are lived out on the ground, rather than how they are intended to function in policy document. The research questions foregrounded in this study are formulated below:

- How do teachers grade students' during CBA?
- To what extent are teachers' grading practices in CBA aligned with the policy mandates?

2. LITERATURE REVIEW

2.1. Grading in formative assessment

In formative assessment, grading is meant to guide instruction and support student progress, not just quantify outcomes. Scholars advocate for criterion-reference approaches that emphasize descriptive feedback and holistic evaluation [6], [9]. Formative grading should also promote fairness and transparency, particularly when students help co-construct assessment criteria [15]. However, in Malaysian education system, these ideals are inconsistently applied. Research indicated English as a second language (ESL) teachers tend to rely on conventional grading practices, such as assigning percentage marks or accumulating task scores in formative settings [16]. Marnizam and Ali [2] affirmed that teachers tend to convert qualitative judgments into numerical evaluations to fulfil administrative reporting requirements. A similar trend was observed in special education contexts [17], where the pressure to generate numerical evidence overshadowed the formative goals of CBA.

Sense-making theory [13] offer insights into this trend. The theory suggests individuals actively interpret policies through their own experiences and institutional contexts. In this study, teachers are seen as policy actors who make sense of CBA expectations in different ways. Some view formative grading as vague or subjective, which raise concerns about fairness and accountability, especially in environments where grading consistency is closely scrutinized. By applying sense-making theory, this study seeks to uncover how teachers interpret and negotiate the meaning of CBA-related grading policies. This discovery can inform their actual grading practices in ESL classroom whether that mean adapting to formative principles, selectively integrating them, or reverting to more traditional summative methods. This lens reveals how cognitive and contextual factors shape the gap between policy and practice.

2.2. Challenges of CBA implementation

While the Ministry of Education has invested in the promotion of CBA to replace outdated exam-oriented paradigms, implementation at the school level has proven difficult. Research across Malaysian contexts indicates a persistent misalignment between reform goals and classroom realities. Several studies [1], [10], [18], [19] revealed that teachers frequently conflate formative assessment with routine monitoring and record-keeping, with limited use of assessment information for diagnostic or instructional decision making. System-level factors such as insufficient professional development, limited institutional readiness, and weak alignment between CBA principles and school management practices further constrain meaningful enactment [10], [19].

From a curriculum reform perspective, these challenges are rooted in deeper tensions between constructivist ideals and entrenched accountability regimes [9], [13], [19]. In many schools, administrative expectations for standardized reporting and clean datasets encourage mark adjustment, data alignment, and compliance-oriented documentation rather than rich formative engagement with learner progress [1]. Collectively, these studies suggest that without coherent system-wide support, spanning teacher education, school culture, and performance metrics, CBA risks being absorbed into existing managerial structures, leaving its transformative potential for teaching and learning largely unrealized.

2.3. Conceptual framework

Guided by sense-making theory [13], this study conceptualizes teachers as active interpreters, not passive recipients of CBA policy. The framework posits a three-stage relationship: i) policy environment, ii) teacher interpretation, and iii) classroom grading practice as shown in Figure 1. First, teachers receive policy signals from official documents, school leadership, training, and administrative expectations. These messages convey what CBA grading should look like in principle (e.g., holistic evidence, formative feedback, and descriptive judgments). Second, teachers interpret these messages through personal beliefs, prior experiences with exam-oriented systems, perceived accountability pressures, and the practical constraints of their school environment. This interpretive process shapes how they understand the purpose, meaning, and feasibility of formative grading. Third, these interpretations directly influence classroom enactment, including decisions about evidence collection, judgment formation, feedback, and whether to adopt formative principles or revert to summative practices. Thus, sense-making theory [13] provides an analytical lens to trace how policy intentions become transformed or distorted through cognitive filtering and contextual pressures, ultimately shaping the alignment (or misalignment) between CBA reform goals and actual ESL grading practices.

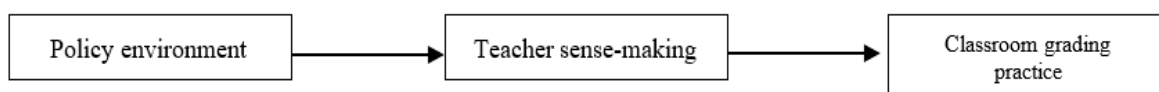


Figure 1. Sense-making theory in CBA grading

3. METHOD

3.1. Research design

This study adopted a qualitative phenomenological approach to explore how ESL teachers interpreted and applied grading practices under the CBA policy. This design was instrumental in gathering in-depth discovery of the participants lived experiences in relation to the phenomenon being studied [20], such as CBA in this case. Six ESL teachers from public secondary schools in Pulau Pinang were selected through purposive sampling with the criterion-based selection. Creswell and Baez [20] suggested that 5 to 25 participants would be sufficient number to meet the purpose of the study. This sampling method was effective in assisting researchers to acquire “information-rich” data tailored to the study’s needs [21]. The participants were in-service educators from Malaysian public secondary schools in Penang. They held a Bachelor’s degree in English Language Teaching (ELT), had at least three years of ESL teaching experience, and attended workshops or training on CBA. Their demographic profiles are shown in Table 1, with pseudonyms used to maintain privacy and confidentiality.

Table 1. Participants’ demographic profile

Participants	Gender	Age	Educational background	Major	Teaching experience (years)
Teacher 1	Male	35	B.Ed. in Teaching ESL	English	7
Teacher 2	Female	28	M.Ed. in Teaching ESL	Linguistics	4
Teacher 3	Female	30	B.Ed. in Teaching English to Speakers of Other Languages	English	6
Teacher 4	Female	27	B.Ed. in Teaching ESL	English	3
Teacher 5	Female	40	M.Ed. in Linguistics	Linguistics	14
Teacher 6	Female	45	B.Ed. in Teaching ESL	English	18

3.2. Data collection methods

The semi-structured interview with open-ended questions served as the primary data collection method. The open-ended format enabled participants to articulate their views freely and in detail, generating rich descriptions of their grading practices in their own words and allowing researchers to uncover nuanced insights that may not surface through structured questioning [3]. An interview guide was employed to ensure consistency across all interview sessions, thereby maintaining alignment with the study’s objectives and reducing interviewer bias [21]. The protocol, adapted from Maguire *et al.* [4], Aziz [14], and Joachim and Hashim [18] who studied curriculum reform in Malaysian primary schools, was modified to explore grading practices in Malaysian ESL classrooms. The protocol included three sections: Section A gathered teachers’ personal background (teaching experience and education). Section B focused on their knowledge of CBA implementation; and Section C explored their grading practices during classroom assessments. Interviews

lasted 20 to 30 minutes, were conducted in English as participants were ELT practitioners, and were recorded for data analysis purposes. The protocol was reviewed by three experts in English education and assessment. Their feedback helped improve the clarity and relevance of the questions. A pilot interview with two ESL teachers (not part of the main study) was conducted to test the clarity and flow of the questions, resulting in minor revisions.

In addition to interviews, document analysis was conducted to triangulate and validate the qualitative findings. Interview data on grading practices were systematically compared with the official CBA policy document [22]. To extract and organize relevant information efficiently, the researchers used a document summary form [21] containing fields such as the document's data, title, purpose, content outline, relevance to the study, and researcher reflections. To ensure orderly referencing, a labelling system (e.g., "document form 1" designated as "doc 1") was implemented to facilitate transparent referencing and ensure traceability of evidence across data sources [3]. As shown in Figure 2, the triangulation process integrated insights from both interviewers and document analysis. Document excerpts served to complement the interview data by either validating teachers' narratives. For example, confirming alignment between teachers' grading practices and policy expectations or highlighting contradictions, such as instances where teachers' interpretations diverged from the official assessment guidelines. The synthesis of these data sources strengthened the interpretation of findings and enabled a more comprehensive understanding of ESL grading practices within the Malaysian CBA context.

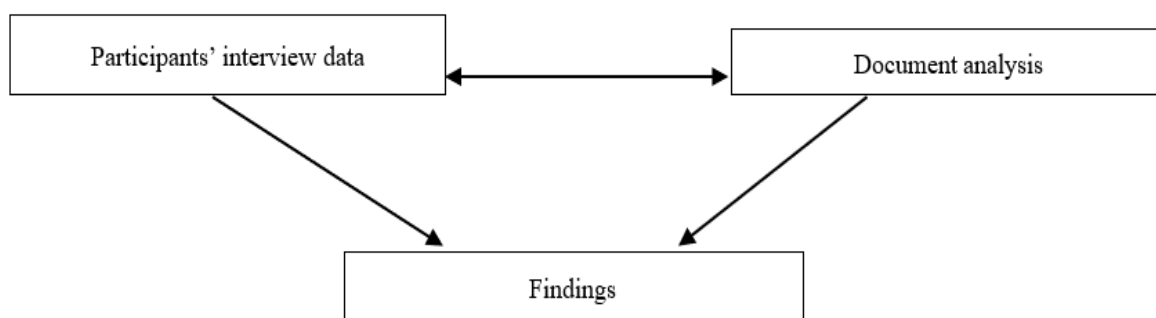


Figure 2. Data triangulation procedures

3.3. Data analysis

This study employed thematic analysis procedures to discover teachers' grading practices in CBA. Figure 3 visually presents the sequential steps of this analytical procedures as suggested by Braun and Clarke [23]. The analysis began with transcribing and rereading interview data using HappyScribe software to ensure familiarity. Meaningful data segments were then identified and assigned descriptive codes using ATLAS.ti, a qualitative data analysis software that helps organize and manage data. This form of qualitative data analysis software assists researchers in storing and organizing their data, enables them to assign labels or codes to their data, and makes it simpler to search the data for texts or words [20]. The purpose of these computer programs is not to generate an analysis, but rather to reduce the researcher's manual workload. Similar codes were grouped to form themes, which were reviewed, refined, and named to capture their essence. Data saturation was reached by the sixth interview, when no new codes or concepts emerged, and the same themes appeared repeatedly across participants' accounts. The recurrence of such themes and quotations across participants as shown in Table 2 confirmed that thematic saturation had been achieved. The final themes were then presented and discussed to reflect the core findings of the study.

To enhance the trustworthiness of the study, several strategies were systematically implemented. Credibility was ensured through methodological triangulation, expert validation of the interview protocol, pilot testing of the instruments, and member checking to verify the accuracy of participants' views. Dependability was strengthened by maintaining detailed documentation of data collection procedures, interview guides, and coding processes, complemented by the principal researcher's formal training in English education and qualitative methods. Confirmability was upheld through an audit trail comprising interview transcripts, document summary forms, analytic memos, and a reflective journal that documented key decisions and minimized researcher bias. Transferability was supported by providing thick descriptions of the research context, participant characteristics, and the CBA policy environment to enable readers to judge the relevance of the findings to similar settings. Additionally, intercoder reliability was established through the involvement

of a secondary coder, and the Cohen's Kappa coefficient of $k=0.646$ indicated a substantial level of agreement, further reinforcing the robustness and trustworthiness of the qualitative analysis.

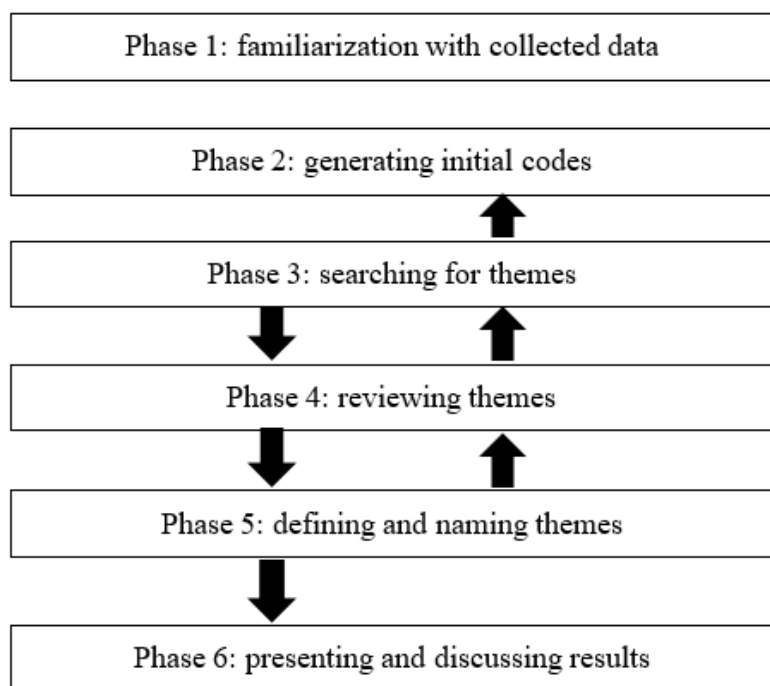


Figure 3. The qualitative process of data analysis

Table 2. Themes development

Example raw data (quotes)	Initial codes	Sub-theme	Major theme
"I always collect their marks in the quizzes... easier for me to track the students' performance." (T3)	Collecting quiz/test marks	Numerical tracking of performance	Numerical-based grading
"...I'll convert students' marks... into the percentage... I can see their level in each language skill." (T2)	Percentage conversion	Use of percentage scores	Numerical-based grading
"...my students were very motivated. It can foster a sense of achievement and competition." (T2)	Motivation through numbers	Motivational value of numerical grades	Numerical-based grading
"We conducted standard setting by presenting the assessment data and comparing it among ourselves." (T1)	Comparing marks in professional learning community (PLC)	Collaborative cut-score decisions	Standard setting
"Teachers meet to set the standards for each group based on the grades obtained by students." (T4)	Group-based grading benchmarks	Shared grading benchmarks	Standard setting
"We have a template provided by the school admin... divided into the specific skills I plan to evaluate." (T3)	School-provided recording template	Documentation for skill-based assessment	The recording template
"In the file, we must include the transit form... two columns—one for the first half of the year and another for the second half." (T2)	Biannual progress tracking	Structured performance monitoring	The recording template
"It requires extensive and laborious effort... during the data entry and preparation phases." (T3)	Time-consuming data entry	Administrative burden of documentation	The recording template
"Sometimes I must adjust the data to align with their requirements." (T3)	Adjusting data for admin compliance	Data modification under pressure	The recording template

4. RESULTS AND DISCUSSION

4.1. Teachers' grading practices

A thematic analysis based on data collected in this study discovered several grading practices that the participants employed. These findings highlight the diverse approaches and pedagogical strategies used to evaluate student performance across different academic settings. Table 3 presents a comprehensive summary of the specific grading practices discovered in this study.

Table 3. Emerging themes and codes representing teachers' grading practices

Major theme	Sub-theme	Initial codes
Numerical-based grading	Numerical tracking of performance; use of percentage scores; motivational value of numerical grades	Collecting quiz/test marks; percentage conversion; motivation through numbers
Standard setting	Collaborative cut-score decisions; shared grading benchmarks	Comparing marks in PLC; group-based grading benchmarks
The recording template	Documentation for skill-based assessment; structured performance monitoring; administrative burden of documentation; data modification under pressure	School-provided recording template; biannual progress tracking; time-consuming data entry; adjusting data for admin compliance

4.1.1. Numerical-based grading

The findings reveal that teachers relied heavily on numerical data to assess students' performance in the formative assessment of CBA. Teachers perceived this approach as an efficient and straightforward method to monitor students' learning progress. For example, Teacher 3 explained that she regularly collected quiz marks to track students' progression after teaching sessions.

"For me, I always collect their marks in the quizzes that I gave them after I teach them. You know, for me it's easier for me to track the students' performance..." (T3)

Teacher 2 further noted that numerical data especially when converted into percentages allowed her to pinpoint student performance levels more precisely across various language skills. She highlighted that numerical grading provided a sense of certainty ("black and white") and helped both teachers and students identify which areas required further improvement.

"...it's easier to see students' performance in the number form. We can have black and white... for me I'll convert students' marks they achieved on the tasks given and converted it into the percentage and this way easier because I can see their level in each language skill...." (T2)

Beyond tracking performance, Teacher 2 also believed that numerical scores had a motivational impact. According to her, students responded positively to high percentage scores, which fostered a sense of achievement and healthy competition.

"...my students were very motivated. It can foster a sense of achievement and competition..." (T2)

However, the numerical-based grading misaligns with formative purposes. The reliance on marks suggests a continuation of traditional, exam-oriented practices that contradict the core principles of CBA. According to the CBA Guidebook [22], classroom assessment should create a low-pressure, student-centered environment that emphasizes learning improvement rather than ranking or comparison.

"Teachers need to create a fun atmosphere, without pressure and modest in the classroom so that the interaction between teachers and students, students and students, as well as students and learning materials can take place in a meaningful situation." [22]

These findings show that numerical grading served multiple roles. Teachers used it to monitor progress, motivate students, and communicate achievement in ways they found clear and manageable. However, the overemphasis on quantification, while practical, may overlook essential learning outcomes such as critical thinking, language use in authentic contexts, and metacognitive development. These findings echo Wu and Yu [10], who found that Malaysia teachers often default to marks and percentages due to familiarity and perceived objectivity, despite policy calls for more qualitative, descriptive forms of classroom assessment.

4.1.2. Standard setting

The grading practices were also conducted through internal standard-setting procedures. It involves aligning student performance with predetermined benchmarks or success criteria, often through collaborative discussions among teachers to foster consistent judgments and reduce grading variability [24]. This is supported by Thirakunkovit [25] who emphasized its role in establishing cut scores and ensuring valid, reliable evaluations. Teacher 1 and Teacher 4 explained that standard setting in their school involved a panel-based moderation process known as PLC in which teachers reviewed students' assessment data collectively to ensure consistency in assigning grades.

“In the panel, we conducted standard setting by presenting the assessment data and comparing it among ourselves. This ensures an even distribution of data across teachers, so there’s no issue of one teacher being too strict while another is too lenient.” (T1)

“Teachers meet to set the standards for each group based on the grades obtained by students. Following these discussions, I recorded my students’ mastery levels based on the agreed standards.” (T4)

These findings suggest that the practice of standard setting, while promoting intra-school consistency, remains largely informal and localized. Teachers’ decisions often stem from professional judgment and group consensus rather than being guided by national policy documents such as Standard-Based English Language Curriculum (SBELC) or Curriculum and Assessment Standard Document (DSKP). This is problematic, as the CBA Guidebook [22] explicitly requires that teachers anchor their assessment decisions in the content and learning standards stipulated in curriculum documents.

“Decision regarding students’ mastery levels should be based on the teacher’s autonomy while adhering to the DSKP’s requirements and specifications.” [22]

Overall, the study uncovered substantial inconsistencies in teachers’ standard-setting practices. Many teachers still use traditional, exam-centric methods that prioritize marks and scores, resulting in inconsistent grading practices across schools. The findings resonate with earlier studies demonstrating the challenges teachers face in aligning assessment decisions with curriculum standards. Wu and Yu [10] found that Malaysian teachers frequently relied on school-level norms and informal consensus when determining mastery levels, particularly when curriculum descriptors were perceived as broad or ambiguous. In broader EFL contexts, Liu [8] observed that teachers tended to default to numerical thresholds and peer norms for standard setting, especially in systems with strong accountability pressures. The present study adds to this literature by showing how the lack of reference to curriculum contributes to inconsistent grading practices, highlighting the need for clearer guidance and stronger curriculum-assessment integration.

4.1.3. The recording templates

This study found that the recording template, or transit form, played a central role in classroom grading. It was designed to document student progress which includes learning objectives and spaces for assessment results. Teachers used it throughout the year to track student competency across assessment cycles. For example, Teacher 2 used a form with two columns—one for each half of the year—allowing for easy comparison of student progress in line with the biannual CBA reporting schedule.

“In the file, we must include the transit form, which contains a list of student names. Since CBA is conducted twice a year, the form must have two columns—one for the first half of the year and another for the second half. I believe the focus is more on monitoring students’ progress throughout the year rather than just on the final results.” (T2)

Similarly, Teacher 3 noted that the template she received from her school was skill-specific and aligned with the curriculum’s content and learning standards. It was used not only for personal tracking but also served as an official document for inspection by school administrators.

“We have a template provided by the school admin. Essentially, the template is divided into the specific skills I plan to evaluate. It’s for my personal records, which I periodically refer to for tracking their progress...” (T3)

While these templates theoretically support the goal of formative assessment by offering a structured way to monitor individual learning trajectories, in practice, they often functioned as tools for institutional oversight rather than pedagogical reflection. Teachers reported that data entry into these templates was time-consuming and burdensome. Teacher 3 expressed frustration at the volume of administrative work associated with completing these forms.

“It requires extensive and laborious effort, especially during the data entry and preparation phases. These demands come from higher authorities...” (T3)

Her account also suggests that the emphasis on record-keeping shifted teachers’ focus from formative instructional strategies to fulfilling documentation requirements. The assessment process became increasingly procedural, prioritizing compliance over meaningful engagement with student learning.

“They consistently critique our data, pointing out parts they consider unreliable while approving others. As a result, there are times when I must adjust the data to align with their requirements...”
(T3)

Teachers are often caught between dual expectations: to provide authentic, student-focused assessments and to meet institutional and administrative accountability targets. From a policy enactment perspective [4], teachers’ use of the recording template reflects a form of fabrication—the production of acceptable performance evidence to satisfy external audits rather than to inform teaching decisions. This suggests a displacement of CBA’s original intent, whereby the process of documenting learning becomes performative rather than developmental. Although the CBA Guidebook [22] advocates for continuous recording to support diagnostic teaching and learner-centered feedback, the implementation realities appear to subvert this aim:

“The recording activity is conducted by teachers periodically to ensure that detailed information about students’ mastery of learning is collected and readily available. This information enables teachers to make professional judgements regarding students’ levels of mastery across a range of Content and Learning Standards.” [22]

Meanwhile, the findings reveal a misalignment between this policy intent and its implementation. Teachers’ judgements were often shaped by institutional pressures rather than curriculum standards, indicating that the recording templates, though designed to support pedagogical decisions, had become tools of managerial control. This disconnect illustrates the challenges of policy translation within systems characterized by top-down accountability. These findings parallel to Liu [8] who found that in many Asian EFL contexts, the pressure to maintain “clean and consistent” assessment records resulted in teachers altering or simplifying data to satisfy institutional audits. Taken together, the interconnectedness of grading practices (numerical-based grading, standard setting, the use of recording templates) suggests a reinforcing loop. Teachers’ reliance on numerical scores feeds into informal standard-setting procedures, where marks are compared and classified without rigorous reference to curriculum document benchmarks. These classifications are then formalized through transit forms, transforming assessment data into administrative artefacts. This cyclical pattern demonstrates how institutionalized quantitative grading can reshape collective practices and obscure the formative intentions of CBA.

5. DISCUSSION

This study identified diverse approaches teachers used in grading students within the CBA context. While these practices offered practicality and order, they also revealed tensions between formative assessment principles and implementation realities. Teachers’ continued reliance on numerical grades reflects the persistent influence of exam-style assessment. Although CBA promotes formative, student-centered evaluation, numbers remain the dominant method, valued for perceived objectivity, clarity, and ease of communication. These traits aligned with high stakes testing culture [10]. Such quantification, while convenient, may overlook critical thinking, authentic language use, and metacognitive skills. Isa *et al.* [19] caution that grades alone cannot capture the full spectrum of learning, advocating more holistic approaches with qualitative feedback, self-assessment, and peer evaluation. Sense-making theory [13] explains this persistence as teachers’ interpretation of reforms through pre-existing mental models shaped by past experience and institutional culture. The reliance on numerical grades and traditional standard-setting reflects a high-stakes mindset, highlighting the gap between policy intent and classroom reality, where numbers offer certainty and accountability but remain ill-suited to formative learning.

While standard-setting procedures through panel discussion encouraged collegiality and aimed to ensure consistency, they were largely anchored in shared professional judgment and collective interpretation of numerical scores rather than curriculum documents. This aligns with Alkin *et al.* [26], who note that teachers often recontextualize policy based on local norms, prior experiences, and institutional pressures. Here, standard setting functioned less to align assessment with curricular standards and more to legitimize traditional grading with the CBA framework. Consequently, despite appearing collaborative and fair, these practices risk perpetuating inconsistencies across schools, undermining the reliability and equity standard setting is meant to ensure [11]. Inconsistencies in cut scores, driven by subjective interpretation of standards, reflect the broader challenge of ensuring equity in CBA implementation. This performative enactment prioritizes producing acceptable evidence of compliance over making pedagogically meaningful judgments.

The use of transit forms, intended to track learning progress, exemplifies how assessment tools can shift from pedagogical purposes to bureaucratic ones. Although they could support formative assessment by

documenting individual growth, teachers reported that these forms served mainly as compliance tools for administrative review rather than professional reflection. This supports Hopfenbeck *et al.* [11] who warns that excessive documentation risks bureaucratizing formative assessment, prioritizing reporting over instructional value. Frustration with data entry and accounts of adjusting data to meet administrative expectations further reveal how accountability pressures can lead to fabricated performance evidence [26].

Overall, these findings depict a self-reinforcing cycle of grading practices within CBA, where entrenched norms of quantification persist. Rather than promoting holistic, process-oriented assessment, the system gravitates toward measures that are easily recorded, compared, and reported which often at the expense of deeper learning. As Abdullahi and Othman [27] cautioned, when assessment is reduced to the measurable, its formative potential is diminished.

5.1. Practical implications

The findings of this study have several implications for stakeholders involved in the implementation of CBA in Malaysia. For policymakers, there is a clear need to reduce bureaucratic burden by streamlining documentation and reporting procedures, allowing teachers to focus on meaningful assessment tasks. Clarifying the formative purpose of CBA and providing annotated exemplars of performance standards across subjects and proficiency levels would support teachers' interpretations of curriculum requirements. Policies should emphasize the use of assessment data for instructional planning and student improvement rather than ranking.

Additionally, as the gap in assessment literacy persist, teacher educators and professional development providers should priorities practical, hands-on learning through case analyses and simulated grading tasks. Training should also strengthen teachers' skills in using performance standards, providing meaningful feedback, and designing alternative formative assessment such as portfolios and reflective journals.

Furthermore, school leaders need to play crucial mediating role in shaping how CBA is enacted. They should foster collaborative decision-making through PLCs and moderation practices that focus on shared understanding rather than compliance. Providing structured support and monitoring while preserving teachers' professional autonomy can encourage more authentic and contextually responsive implementation of CBA.

Teacher is also encouraged to shift from assessment for comparison to assessment for learning. This requires using curriculum standards as central tools for instructional planning a formative decision making. Integrating reflective and student-centered tools such as learning logs, checklist, self-assessment rubrics, and descriptive feedback can promote students' metacognitive awareness and continuous development. Purposeful engagement with performance standards will help ensure grading decisions align with the formative aims of CBA.

6. CONCLUSION

In conclusion, this study has examined how Malaysian ESL teacher interpret and enact grading practices within the CBA system. Although teachers demonstrated strong commitment to supporting student learning, their grading approaches remained largely influenced by long-standing exam-oriented traditions. Numerical scores continued to dominate assessment decisions due to their perceived objectivity, ease of communication, and alignment with institutional expectations. These practices, however, run counter to the formative aims of CBA, which emphasize descriptive judgments, continuous feedback, and holistic evaluation of learner development. Standard-setting procedures, while intended to ensure consistency, also reflected localized interpretations shaped by peer norms rather than curriculum-anchored performance standards. Teachers relied heavily on informal consensus during PLC meetings, often using numerical thresholds instead of criterion-referenced descriptors. Similarly, recording templates designed to guide formative documentation were frequently perceived as compliance tools, resulting in administrative burdens that constrained teachers' ability to use assessment data meaningfully for instructional planning. Collectively, these findings reveal a tension between policy ideal and classroom realities. Sense-making theory helps explain this gap by illustrating how teachers interpret reform messages through the lens of past experiences, institutional accountability, and practices classroom constraints. The result is a reform that appears structurally implemented but pedagogically diluted.

Moreover, this study has several limitations worth noting. Firstly, because it explores teachers' grading practices in CBA using a qualitative design, it does not evaluate the actual effectiveness of these practices in improving student learning. Future research could therefore employ quantitative or mixed-methods approaches such as quasi-experiments, large-scale surveys, or correlational studies to examine how different grading methods influence students' language development, motivation, and classroom engagement. Secondly, this study focuses primarily on teachers' perspectives, which limits the understanding of how other stakeholders experience and interpret CBA reform. Future studies should incorporate the views of students, parents, and school administrators to capture how grading practices shape learning experiences and expectations across the broader educational ecosystem. Finally, given that teachers' grading behaviors may evolve as they become more familiar with CBA, longitudinal research is recommended to track how teachers

adapt their grading practices over time and how sustained exposure to policy, training, and school-level support influences their implementation of formative assessment principles.

ACKNOWLEDGMENTS

The researchers would like to express sincere thanks to Faculty of Education, Universiti Malaya, Kuala Lumpur, Malaysia for the invaluable support and resources provided throughout the course of this research.

FUNDING INFORMATION

The authors declare that no funding was received for the research, authorship, and/or publication of this article.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Anwar Farhan Mohd Marzaini	✓	✓	✓		✓	✓	✓	✓	✓			✓	✓	
Shahazwan Mat Yusoff	✓	✓				✓		✓	✓			✓		
Muhammad Usamah Mohd Ridzuan	✓	✓		✓	✓	✓	✓			✓			✓	
Nur Fatin Shahmina Mohd Fauzey	✓	✓	✓		✓	✓	✓		✓	✓	✓			
Nur Husna Serip Mohamad	✓	✓	✓		✓		✓		✓	✓				
Noraziah Mohd Amin	✓	✓	✓		✓	✓	✓			✓				
Muhammad Aiman Abdul Halim	✓	✓	✓		✓	✓	✓		✓	✓	✓			
Syafiq Munir Ismail Munir	✓	✓			✓	✓	✓		✓	✓	✓			

C : **C**onceptualization

M : **M**ethodology

So : **S**oftware

Va : **V**alidation

Fo : **F**ormal analysis

I : **I**nvestigation

R : **R**esources

D : **D**ata Curation

O : Writing - **O**riginal Draft

E : Writing - Review & **E**editing

Vi : **V**isualization

Su : **S**upervision

P : **P**roject administration

Fu : **F**unding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the Ministry of Education Malaysia [KPM.600-3/2/3-eras (12625)].

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, [AFMZ]. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

- [1] A. F. M. Marzaini, S. M. Yusoff, N. Zakaria, M. H. Basal, W. N. E. H. Sharil, and K. Supramaniam, "A discovery on Malaysian ESL teachers' adherence to policy mandates in classroom-based assessment practices," *Malaysian Journal of Learning and Instruction*, vol. 21, no. 1, pp. 249–284, Jan. 2024, doi: 10.32890/mjli2024.21.1.9.
- [2] F. I. Marnizam and S. R. Ali, "Evaluation of the implementation of classroom assessment (PBD) among primary school mathematics teachers," *Jurnal Pendidikan Sains dan Matematik Malaysia*, vol. 11, no. 2, pp. 81–94, May 2021, doi: 10.37134/jpsmm.vol11.2.7.2021.
- [3] J. Pilarska, "The constructivist paradigm and phenomenological qualitative research design," in *Research Paradigm Considerations for Emerging Scholars*, Channel View Publications, 2021, pp. 64–83, doi: 10.2307/jj.22730542.10.
- [4] M. Maguire, S. Gewirtz, E. Towers, and E. Neumann, "Contextualising policy work: policy enactment and the specificities of English secondary schools," *Research Papers in Education*, vol. 35, no. 4, pp. 488–509, Apr. 2019, doi: 10.1080/02671522.2019.1601758.
- [5] H. F. Lokman, N. Abdullah, M. C. Mustafa, S. Sembak, and V. L. Baskaran, "Metasynthesis of classroom-based assessment implementation: impact and issues," *International Journal of Learning, Teaching and Educational Research*, vol. 22, no. 4, pp. 310–325, Apr. 2023, doi: 10.26803/ijlter.22.4.18.
- [6] L. B. Nilson, "Essentials of specifications grading," in *Specifications Grading*, Routledge, 2023, pp. 56–66, doi: 10.4324/9781003447061-5.
- [7] L. Ketonen and J. H. Nieminen, "Professional autonomy vs. assessment criteria: teacher agency in the midst of assessment reform," *Journal of Curriculum Studies*, vol. 58, no. 2, pp. 257–274, Feb. 2025, doi: 10.1080/00220272.2025.2460476.
- [8] Y. Liu, "The effect of classroom-based assessment (CBA) on Chinese EFL learners' speaking performance, engagement, and willingness to communicate: Undergraduate students' perceptions," *Heliyon*, vol. 10, no. 12, p. e32629, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32629.
- [9] N. F. Mohd Salleh, "Washback effect of school-based assessment (SBA) on Malaysia secondary school students' English language learning," *PASAA*, vol. 66, no. 1, pp. 168–201, Oct. 2023, doi: 10.58837/chula.pasaa.66.1.5.
- [10] J. Wu and X. Yu, "The influence of formative assessment on academic performance: exploring the role of teachers' emotional support," *Frontiers in Psychology*, vol. 16, Apr. 2025, doi: 10.3389/fpsyg.2025.1567615.
- [11] T. N. Hopfenbeck, Z. Zhang, S. Z. Sun, P. Robertson, and J. A. McGrane, "Challenges and opportunities for classroom-based formative assessment and AI: a perspective article," *Frontiers in Education*, vol. 8, Nov. 2023, doi: 10.3389/educ.2023.1270700.
- [12] A. Akbari and M. Shahrokhi, "Unveiling fairness in scoring: a thorough method for precise cutoff score calculation in education assessment," *Quality Assurance in Education*, vol. 32, no. 3, pp. 493–509, Jun. 2024, doi: 10.1108/qaec-12-2023-0208.
- [13] D. O. Braithwaite and P. Schrodt, *Engaging theories in interpersonal communication: multiple perspectives*. Routledge, 2021, doi: 10.4324/9781003195511.
- [14] A. A. Aziz, "An investigation of the relationship between curriculum policy and assessment practice in Malaysian classrooms," PhD thesis: University of Glasgow, 2021.
- [15] A. D. Bazvand and A. Rasooli, "Students' experiences of fairness in summative assessment: a study in a higher education context," *Studies in Educational Evaluation*, vol. 72, p. 101118, Mar. 2022, doi: 10.1016/j.stueduc.2021.101118.
- [16] O. Ketsman and T. Reeves, "K-12 teachers' implementation of standards-based grading in the classroom: a quantitative study," *Journal of School Administration Research and Development*, vol. 8, no. 2, pp. 65–75, Oct. 2023, doi: 10.32674/jsard.v8i2.5565.
- [17] Y. Ramma, A. Bhola, and M. Watts, "In-service physics teachers' content knowledge: a critical reflection on the case of the upthrust concept," *Education Inquiry*, pp. 1–31, Oct. 2024, doi: 10.1080/20004508.2024.2412878.
- [18] A. Joachim and H. Hashim, "ESL teacher's knowledge and readiness on the implementation of school-based assessment (SBA) in Malaysian primary school," *Creative Education*, vol. 12, no. 05, pp. 1066–1078, 2021, doi: 10.4236/ce.2021.125079.
- [19] A. M. Isa, A.-A. Mydin, and A. G. K. Abdullah, "School-based management (SBM) practices in Malaysia: a systematic literature review," *International Journal of Academic Research in Business and Social Sciences*, vol. 10, no. 9, Sep. 2020, doi: 10.6007/ijarbs/v10-i9/7870.
- [20] J. Creswell and J. C. Báez, *30 essential skills for the qualitative researcher, 2nd ed.* vol. 37. Thousand Oaks, CA: SAGE Publications, 2020.
- [21] M. Ahmad and S. Wilkins, "Purposive sampling in qualitative research: a framework for the entire journey," *Quality & Quantity*, vol. 59, no. 2, pp. 1461–1479, Dec. 2024, doi: 10.1007/s11135-024-02022-5.
- [22] Ministry of Education Malaysia, *Guidelines for classroom assessment, 2nd ed.* (in Malay) Curriculum Development Division, 2019.
- [23] V. Braun and V. Clarke, *Thematic analysis: a practical guide*. Thousand Oaks, California, AS: SAGE Publications, 2021.
- [24] L. Adie, C. Wyatt-Smith, M. Finch, and C. DeLuca, "Evidence of teacher assessment work and its relationship to their assessment identity," *Teaching and Teacher Education*, vol. 141, pp. 104518–104518, Apr. 2024, doi: 10.1016/j.tate.2024.104518.
- [25] S. Thirakunkovit, "Establishing a fair cut score for an in-house English test: a case study on integrating two standard-setting methods," *Language Testing in Asia*, vol. 15, no. 1, Mar. 2025, doi: 10.1186/s40468-025-00348-x.
- [26] M. C. Alkin, A. T. Vo, and C. A. Christie, *Evaluation essentials*. New York City: Guilford Publications, 2024.
- [27] M. Abdullahi and N. Othman, "Bridging the gap between policy intent and implementation," *Journal of Science, Technology and Innovation Policy*, vol. 6, no. 1, pp. 24–33, 2020, doi: 10.11113/jostip.v6n1.49.

BIOGRAPHIES OF AUTHORS



Anwar Farhan Mohamad Marzaini    is a senior lecturer at the Department of Curriculum and Instructional Technology, Faculty of Education, Universiti Malaya, 50603, Kuala Lumpur, Malaysia. He holds a PhD in Education, specializing in Language Assessment. Formerly an ESL teacher for five years, he also served as an SPM examiner for three consecutive years. His research interests include learning-oriented assessment, teacher training, curriculum and instruction, and teaching pedagogy in English language. He can be contacted at email: are_not46@yahoo.com.



Shahazwan Mat Yusoff    is a senior lecturer at the Department of Curriculum and Instructional Technology, Faculty of Education, Universiti Malaya. He holds a PhD in Education (Curriculum and Instruction) from UM, an M.Ed. in TESL, and a B.Ed. (Hons.) in Physical and Health Education from UiTM. With extensive experience teaching Physical Education and English in secondary schools, he has also served as a national English SPM examiner. His research spans assessment, curriculum development, inclusive education, physical and health education, and ESL. He has published widely, presented at international conferences, and reviewed for academic journals. He can be contacted at email: shahazwan@um.edu.my.



Muhammad Usamah Mohd Ridzuan    is a lecturer at the Academy of Language Studies, Universiti Teknologi MARA, Pulau Pinang, Malaysia, specializing in Language Policy and Language Mediation. His primary research area is in the Didactics of Foreign Languages, where he focuses on approaches and techniques for teaching foreign languages. Additionally, he is a certified DELF Examiner (Diplôme d'Études en Langue Française), highlighting his professional qualifications in assessing French language proficiency. He can be contacted at email: usamah@uitm.edu.my.



Nur Fatin Shamina Mohd Fauzey    is a lecturer at the Academy of Language Studies, Universiti Teknologi MARA, Pulau Pinang. She holds a B.A. (Hons.) in English for Professional Communication and an M.A. in Linguistics and English Language Studies. Her research focuses on linguistics, language systems, and applied communication. She can be contacted at email: fatinshamina@uitm.edu.my.



Nur Husna Serip Mohamad    is a senior lecturer at the Academy of Language Studies, Universiti Teknologi MARA, Pulau Pinang. She holds a PhD in English (Sociolinguistics) from the University of Liverpool. Her research focuses on online ethnography, language variation, discourse analysis, and English varieties, with a particular interest in Manglish, intercultural discourse, and identity. She has presented and published internationally and is involved in curriculum design and language teaching innovation. She can be contacted at email: nurhusna@uitm.edu.my.



Noraziah Mohd Amin    is a senior English lecturer at the Academy of Language Studies, Universiti Teknologi MARA Penang Branch. She holds a Bachelor's degree in English Language and Literature, a Master's degree in Applied Linguistics, and a PhD in Communication. With over 16 years of teaching experience, her research and professional interests include English language teaching and learning, applied linguistics, and marketing communication, particularly brand slogans. She is an active researcher and writer, contributing regularly to academic journals and local and international magazines. Her publications include articles in Scopus, Web of Science, and MyCite-indexed journals, conference proceedings, book chapters, and creative works published in local and international magazines and bulletins. She can be contacted at email: noraziahm@uitm.edu.my.



Muhammad Aiman Abdul Halim    is a lecturer at Universiti Teknologi MARA, Penang Branch. He holds a B.Ed. in teaching ESL from UiTM and an M.Ed. in Instructional Technology from UM. His expertise includes teaching ESL, instructional design, and instructional technology, with research interests in e-learning, the technology acceptance model, and knowledge construction. He can be contacted at email: muhd_aiman@uitm.edu.my.



Syafiq Munir Ismail Munir    is the Head of Biodiversity and Conservation at Malaysian Animal Association and Teaching Assistant at Universiti Malaya. He holds a PhD in Science and an M.Sc. in Sustainability Science from the Universiti Malaya, as well as a B.Sc. (Hons.) in Environmental Science (Biodiversity and Conservation) and a Diploma in Environmental Health from UNISEL. His expertise covers biodiversity conservation, environmental sustainability, and education science curriculum. He can be contacted at email: inhsyafiq@gmail.com.