

From Formative Policy to Performative Grading Practices: ESL Teachers' Enactment of Classroom-Based Assessment in Malaysian Primary Schools

Anwar Farhan Mohamad Marzaini^a and Shahazwan Mat Yusoff^{b*}

^a Department of Curriculum & Instructional Technology, Faculty of Education, Universiti Malaya, Kuala Lumpur, Malaysia

^b Department of Curriculum & Instructional Technology, Faculty of Education, Universiti Malaya, Kuala Lumpur, Malaysia

*Corresponding author: shahazwan@um.edu.my

Article information

Abstract

Classroom-based assessment (CBA) is positioned as a cornerstone of Malaysia's assessment reform. Yet evidence of how it is enacted in English as a Second Language (ESL) classrooms remains limited. This qualitative case study examines ESL teachers' grading practices and evaluates their alignment with expectations articulated in national CBA guidance and curriculum standards. Data were generated through semi-structured interviews with six ESL teachers who taught in three primary public schools in Pulau Pinang, Malaysia. Relevant policy and curriculum documents and school recording artefacts were analysed to triangulate teachers' accounts. Thematic analysis identified three dominant patterns. First, teachers relied heavily on numerical mark accumulation and percentage conversion to judge progress and motivate learners. Second, grading decisions were shaped through local standard-setting activities within panels and professional learning communities. These processes often produced cut scores that were not explicitly anchored to national performance descriptors. Third, teachers documented student evidence and recorded mastery using templates such as transit forms. Documentation supported professional judgment for some teachers. For others, it functioned mainly as administrative compliance and increased workload. Overall, teachers reported structured routines and multiple assessment strategies. However, grading enactment frequently reflected exam-oriented logics that conflicted with the intended formative emphasis of CBA. The study highlights a policy–practice gap that may weaken the credibility and comparability of mastery judgments. It recommends clearer exemplars for standards-referenced decisions and sustained professional learning that focuses on interpreting evidence over time and using assessment information to guide feedback and intervention in routine practice.

Keywords

Classroom Based Assessment, Formative Assessment, Grading Practices, ESL Teachers

APA citation:	Marzaini, A. F. M. & Yusoff, S. M. (2026). From formative policy to performative grading practices: ESL teachers' enactment of classroom-based assessment in Malaysian primary schools. <i>PASAA</i> , 74, 24–50.
----------------------	---

1. Introduction

Exam-centric cultures have been a subject of criticism in Malaysian education for a considerable amount of time (Mohamad Marzaini et al., 2024). Diverse stakeholders have expressed disagreement with this approach to assessing students' progress in learning, asserting that it is insufficient for a comprehensive evaluation of such progress. Academics such as Chin et al. (2019) have highlighted that this culture preoccupied with examinations often neglects the critical function of teachers in promoting successful instruction and learning, thus impeding students' progress toward reaching their utmost capabilities. The excessive emphasis on high-stakes examinations at primary level such as *Ujian Pencapaian Sekolah Rendah* (UPSR) has been argued to contribute to inequality by disproportionately affecting "low performers" (Machin et al., 2020). UPSR was a national primary school examination (typically taken at the end of Year 6). Historically, UPSR functioned as a gatekeeping mechanism for academic placement and progression, reinforcing exam-oriented schooling cultures.

Hence, in response to this criticism, the Ministry of Education commenced a reform initiative to establish a progressive assessment system that departs from cultures characterised by excessive testing. Consequently, substantial modifications were implemented in 2021, including the elimination of high-stakes examinations like the UPSR. Replacing UPSR, the government implemented Classroom-Based Assessment (CBA). By incorporating formative and summative assessment methods, the CBA aims to usher in a new testing culture by facilitating a more comprehensive evaluation of students' progress and achievements. By promoting a more positive learning environment, the integration of CBA into the educational system is intended to alleviate students' undue stress and pressure (Mat Yusoff et al., 2025a; Mohamad Marzaini et al., 2023).

The reform has brought significant changes in the education system. The emphasis on formative medium as an assessment method has changed the roles of teachers in the teaching and learning process. Thus, teachers' responsibilities in CBA span all stages, including planning, implementation, assessment, evaluation, reporting, and taking follow-up actions in accordance with guidelines set by policymakers (Mohd Isa et al., 2022). This implies that teachers wield complete control in the assessment process, actively participating in gauging students' progress during instruction.

While teachers were expected to be adept at handling CBA, several studies reveal instances where they struggle to adapt, leading to contentious issues in policy planning and implementation (Thompson, 2018). Crucially, when teachers' classroom grading practices diverge from policy expectations, the reform risks becoming symbolic rather than functional. That said, that is where CBA exists "in name," but daily assessment remains exam-driven. Such misalignments are problematic because they can (i) undermine the intended formative purpose of CBA, (ii) produce inconsistent mastery judgments across teachers and schools, (iii) weaken the credibility and fairness of reporting to parents and stakeholders, and (iv) reduce teachers' capacity to use assessment evidence to adjust instruction and support learning progression (Yusoff et al., 2025). For instance, Mohamad Marzaini et al. (2023) discovered a significant lack of literacy in executing classroom assessments among ESL teachers in the

Pulau Pinang region. Arumugham (2020) emphasised that teachers continue to need consistent supervision and support to fully understand and implement CBA as required. This highlights the challenges educators encounter when incorporating new methodologies into real classroom environments.

Moreover, this shift was claimed to be notably drastic compared to previous changes. Officially introduced by the Ministry of Education in October 2018, implementation at the school level commenced in January 2019. Over these four months, the CBA framework was disseminated to school administrators, and it is imperative to convey this information to teachers and stakeholders directly involved (Marnizam & Ali, 2021; Mat Yusoff et al., 2025b). Borman et al. (2003) suggest that the maturation of curriculum reform is typically strongest between 8 to 14 years after initiation, prompting a query whether teachers possess the competence to implement CBA or if this competence will develop as they begin implementing CBA. Ghavifekr et al. (2016) further lamented that this quick transition of assessment climate coupled with the limited training and dissemination phase has inadvertently caused great resistance among teachers to adapt to the change. As a result, the expectation that teachers will fully master the new assessment policy remains uncertain.

While the research on CBA shows promise nationwide, a thorough examination of previous literature indicates that discussions have predominantly focused on teachers' assessment literacy (Abdul Razak et al., 2023; Marnizam & Ali, 2021) and teachers' readiness (Hanapi et al., 2022; Mohd Faudzi & Ahmad, 2022) in the implementation of new assessment reform. Studies examining how teachers enact grading practices within CBA remain relatively limited, particularly in relation to the alignment between classroom practices and mandated assessment policy. Even though some studies explore teachers' assessment strategies and methods during the implementation of CBA (Arumugham, 2020; Khory et al., 2021; Primus & Mosin, 2021), discussion on teachers' grading practices in primary school is scarce, especially on the alignment between teachers' grading practices and the mandated policy. Therefore, the present study intends to address this gap by exploring the approaches primary school teachers used in grading CBA and assessing how their practices align with the policy requirements. The findings are expected to benefit policymakers and curriculum/assessment divisions, as this study could inform the identification of policy–practice gaps that require clearer guidance and support both the school administrators and instructional leaders by informing targeted professional development and monitoring practices, and ESL practitioners by offering grounded insights into feasible grading approaches that better align classroom assessment with CBA's formative intent. The research objectives and research questions foregrounded in this study are formulated below:

Research Objectives

- 1) To examine how primary school ESL teachers enact grading practices within Classroom-Based Assessment.
- 2) To investigate how teachers' grading decisions reflect, interpret, and negotiate policy expectations in Classroom-Based Assessment.

Research Questions

- 1) How do primary school ESL teachers enact grading practices during Classroom-Based Assessment?
- 2) How are policy expectations interpreted and reflected in ESL teachers' grading practices within Classroom-Based Assessment?

2. Literature Review

2.1 The Classroom-Based Assessment

The primary objective behind implementing Classroom-Based Assessment (CBA) was to nuance the examination-oriented culture that had previously prevailed. By actively evaluating students' skill levels following the curriculum document, CBA adheres to the fundamental tenets of School Assessment. Considered a continuous procedure seamlessly incorporated into the pedagogy and learning environment, this assessment system offers invaluable student progress insights. Moreover, this evaluation is conducted utilising both formative and summative methods, which numerous scholars consider to constitute a comprehensive assessment framework.

As teachers integrate formative and summative assessments in the classroom, the relationship between these two types of assessments is better understood as a continuum rather than a strict dichotomy (Mat Yusoff et al., 2024). According to Black and Wiliam (1998), instructional learning and formative assessment are integral components of a lesson plan. Additionally, formative assessment aligns with Vygotsky's Zone of Proximal Development, enabling instructors to guide students toward higher levels of achievement beyond their current capabilities (Black, 2015; Eli & Yamat, 2021). Classroom assessment is now recognised as an ongoing, curriculum-integrated process.

In contrast, summative assessment serves as an accountability measure, typically used for grading purposes, evaluating whether students have reached a specified level of learning. This evaluation can occur at the end of units, chapters, terms, or the academic year (Eli & Yamat, 2021). Drawing on contemporary learning theories, classroom assessments are often categorized into three types: Assessment of Learning, Assessment as Learning, and Assessment for Learning (Eli & Yamat, 2021). Malaysian classrooms follow the same principles.

“Assessment of learning” signifies summative assessment practices where teachers assess pupils' achievements against defined benchmarks and prepare for future lessons based on outcomes. In comparison, “assessment as learning” captures students' participation in tracking their progress, promoting self-reflection and self-regulation through prompts. Meanwhile, “assessment for learning” involves teachers providing constructive feedback through observations, anecdotal records, interactive questioning, and simple tests to motivate students to improve their learning process.

Given this framing, the application of CBA requires educators to professionally self-govern in planning, designing assessment tasks, administering them, and evaluating students' mastery levels. A flow chart depicting the steps in CBA execution in classroom instruction is presented in Figure 1.

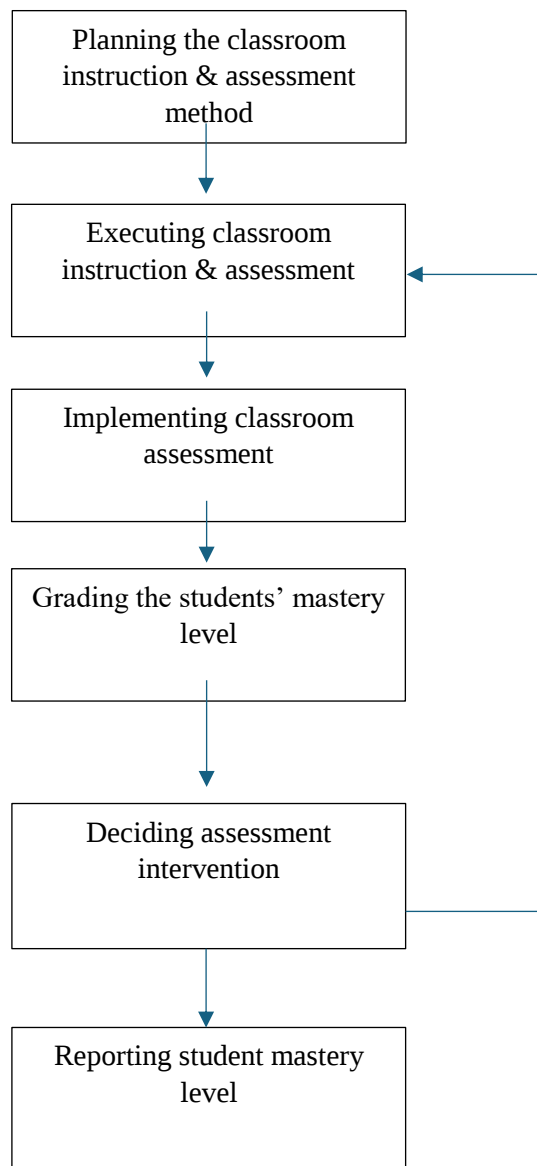
Figure 1*Flow Chart of CBA Implementation*

Figure 1 illustrates the phases of CBA. The process begins with planning instruction and assessment. At this stage, teachers identify learning objectives and ensure that assessment methods align with curriculum standards (Ministry of Education Malaysia, 2019). They also determine suitable assessment components to measure students' progress. Next is the implementation phase, where assessment is integrated into daily teaching. Teachers evaluate students' mastery by comparing their performance with learning standards and benchmarks. Rubrics are often used to record results systematically, allowing teachers to analyse performance and improve instruction. Continuous assessment, or assessment for learning, provides immediate and constructive feedback to support student improvement (Patallion Jr et al., 2016; Tee & Ahmed, 2014). Teachers may also carry out follow-up interventions, either immediately or through planned strategies, to address different learning needs. Since students have varying abilities, assessment helps teachers identify those who have fully or partially understood the content. This enables targeted support to enhance students' proficiency. Finally, teachers prepare progress reports to inform parents about students' knowledge, skills, values,

attitudes, and overall development, helping parents monitor their child's learning more effectively.

2.1.1 Policy foundation: The CBA Guidebook (2019)

Because CBA is guided by policy documents that specify how teachers are expected to plan, assess, record, and report learning, it is important to foreground the Ministry of Education's CBA Guidebook (Ministry of Education Malaysia, 2019) as a key reference point for implementation. The guidebook outlines the intended formative emphasis of CBA, including the use of assessment evidence to support learning progression, inform instructional adjustments, and plan follow-up interventions rather than to rank or compare students. It also provides practical guidance on recording and reporting students' mastery levels and underscores the role of teachers' professional judgement in determining mastery decisions based on sufficient evidence collected over time. In the present study, the CBA Guidebook (Ministry of Education Malaysia, 2019) served as an analytic benchmark in the document analysis to evaluate the alignment between teachers' reported grading practices and policy expectations.

2.1.2 SBELC performance standards and mastery-level decisions

In the Malaysian ESL context, the CEFR-aligned Standard-Based English Language Curriculum (SBELC) provides performance standards that describe expected proficiency and mastery descriptors used to support criterion-referenced judgement. These performance standards serve as reference points for assessing students' mastery of curriculum expectations rather than peer performance (Mohamad Marzaini et al., 2024). Accordingly, CBA grading is expected to draw on SBELC performance standards (together with content and learning standards in the DSKP) when teachers determine mastery levels, record progress, and report achievement.

2.2 The Grading Practices

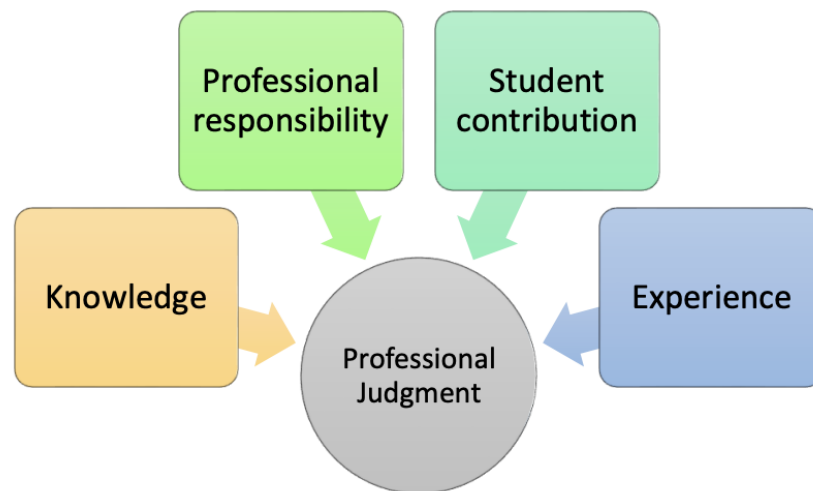
Teachers' grading practices play a pivotal role in ensuring the efficacy of reporting students' achievements. Teachers must consider a spectrum of factors that may influence their grading decisions. Determining students' achievement requires an awareness of various elements that could impact their grades. Additionally, teachers' educational principles, values, and beliefs about teaching and learning are integral to grading decision-making (Sun & Cheng, 2014).

Assessment and grading practices are deemed complex processes (Mohamad Marzaini & Mat Yusoff, 2026; Widiastuti, 2018). Numerous researchers have delved into classroom assessment and grading practices, revealing that grading decisions are intricate and influenced by internal and external factors (Brookhart & Durkin, 2003; McMillan & Workman, 1998). In this study, "internal factors" refer to teacher-level influences such as teachers' beliefs about learning and assessment, pedagogical content knowledge, assessment literacy, prior experiences with exam-oriented systems, and professional judgement when interpreting student evidence. "External factors" refer to contextual and institutional influences such as school assessment policies, administrative monitoring requirements, documentation expectations, PLC/panel standard-setting norms, parental expectations, and broader examination-oriented cultures that can shape grading decisions even under a formative policy

framework. Assessment discrepancies have led many teachers to perceive assessment and grading practices as challenging. Consequently, teachers must employ critical and logical thinking during grading practices to accommodate the myriad factors at play (Widiastuti, 2018). Hence, to grade students' performance, this practice stands on four main elements that a teacher must have to assess their students, as depicted in Figure 2 below.

Figure 2

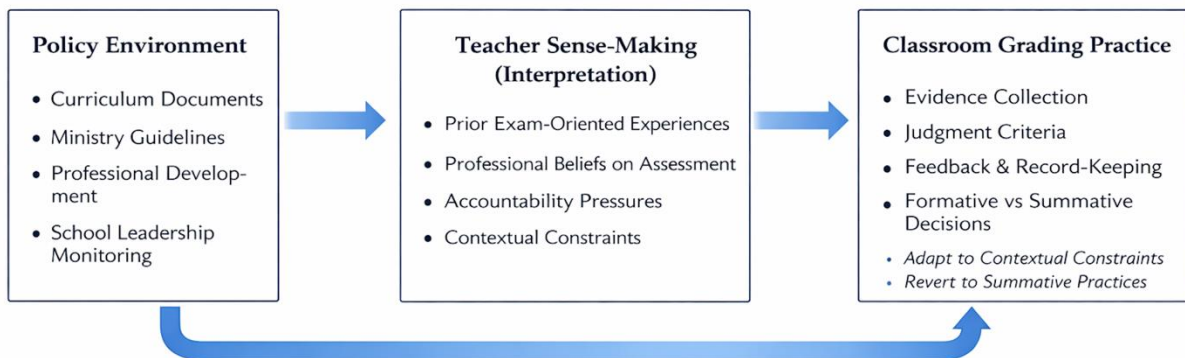
Elements of Professional Judgment (Ministry of Education Malaysia, 2019)



Allal (2013) defined professional judgment as the knowledge, experience, and ethical responsibility teachers use when making instructional decisions. It involves evaluating students' learning and deciding on teaching methods, classroom management, and assessment practices (Ministry of Education Malaysia, 2019). Research shows that professional judgment is shaped by teachers' pedagogical content knowledge—their understanding of subject matter, teaching methods, and students' needs (Abdul Razak et al., 2023; Mat Yusoff et al., 2023). Mohamad Marzaini and Mat Yusoff (2026) highlight four key elements: knowledge, professional responsibility, student input, and experience. In CBA, professional judgment is essential because teachers must use different types of evidence—such as assignments, quizzes, projects, participation, and observations to evaluate learning. It requires interpreting student work beyond scores, adapting grading based on progress, and ensuring fairness and consistency (Mat Yusoff et al., 2025a; Thiede et al., 2018). Given the challenges teachers face in balancing beliefs, policies, and grading demands, studying assessment and grading practices is important to improve implementation across primary school levels.

2.3 Conceptual Framework

Guided by Sense-Making Theory (Hay, 2024), this study conceptualises teachers as active agents who interpret, reconstruct, and enact policy rather than merely implementing it mechanically. Within the context of CBA, grading practices are therefore understood as the product of teachers' interpretive processes situated within institutional and policy environments. Figure 3 below illustrates the conceptual framework of the study.

Figure 3*Sense-Making Theory in CBA Grading*

First, teachers encounter policy signals transmitted through official curriculum documents, ministry guidelines, professional development programmes, school leadership directives, and administrative monitoring mechanisms. These sources communicate normative expectations regarding CBA grading. For instance, emphasising holistic evidence, formative feedback, continuous assessment, and descriptive judgment over numerical scoring.

Second, teachers engage in sense-making. They interpret policy messages through the lens of their prior experiences (particularly within exam-oriented systems), professional beliefs about assessment, perceptions of accountability and stakeholder expectations, and contextual constraints such as workload, time limitations, and class size. This interpretive phase is not neutral; it is shaped by cognitive schemas and institutional pressures that influence how teachers understand the purpose, legitimacy, and practicality of formative grading reforms.

Third, these interpretations are translated into classroom enactment. Teachers' grading decisions, including evidence selection, judgment criteria, feedback practices, record-keeping methods, and the balance between formative and summative elements, reflect their reconstructed understanding of CBA policy. Where tensions arise between reform ideals and contextual realities, teachers may adapt, modify, or revert to more traditional summative grading approaches.

Thus, Sense-Making Theory provides a coherent analytical lens for tracing how policy intentions are filtered through teacher cognition and contextual conditions, resulting in varying degrees of alignment or misalignment between CBA reform goals and actual ESL classroom grading practices. The framework therefore positions grading not as a technical procedure, but as a socially situated and cognitively mediated practice.

3. Methodology

3.1 Research Design

In this study, a qualitative approach was chosen as the foundation method to gather data related to primary school ESL teachers' grading practices in CBA. The study employed a case study design (Yin, 2014), as it allowed a more comprehensive exploration of teachers' classroom-based experiences and practical enactment of CBA at the grassroots level. Merriam and Tisdell (2015) asserted that the case study design can be instrumental in gathering in-depth discovery in relation to the phenomenon being studied. Furthermore, Rahman (2014) noted that qualitative case study approaches are useful for examining how participants interpret and respond to curriculum reform in real educational settings. In the present study, this approach

enabled an in-depth investigation of ESL teachers' experiences, interpretations, and enactment of CBA grading practices within their school contexts. This study aims to capture the nuances of these policies through the lenses of practitioners by assessing teachers' experiences with the current evaluative system in place to understand the workings, issues, and triumphs of policy execution. Fundamentally, the research employed a case study design as an effective method to thoroughly investigate primary school ESL teachers' practical implementation of CBA. This approach revealed the superficial mechanisms and the more profound situational elements influencing its execution. This methodology was supported by substantial scholarly literature, adhering to established qualitative research principles and the imperative to understand the complex and ever-changing aspects of educational policy execution.

3.2 Participants and Sampling Techniques

This research was conducted in the public primary schools around Pulau Pinang state, Malaysia. The study focused on public primary schools because CBA implementation is governed by the national Ministry of Education policy and is most uniformly enacted within the public-school system, where teachers follow the same mandated curriculum standards, reporting structures, and administrative monitoring procedures. Restricting the sample to public primary schools also strengthened contextual comparability across participants (e.g., similar assessment policy expectations, school accountability structures, and shared examination-oriented pressures), thereby enabling a clearer interpretation of grading practices under the same policy environment (Yusoff et al., 2025). Private schools and other educational levels may operate under different assessment expectations, curricular structures, and stakeholder pressures; therefore, they were excluded to maintain a coherent case boundary for this qualitative inquiry.

Six ESL teachers were selected as participants in this study. The purposive sampling technique (Merriam & Tisdell, 2015) was employed in choosing the samples out of the larger population. This sampling method is suitable in the context of the present study, as Merriam and Tisdell (2015) affirmed for its effectiveness in acquiring "information-rich" data tailored to the study's needs. These ESL teachers were also selected based on the criterion-based selection method. Hence, the ESL teachers who were included in this study were teaching within-service teachers at public primary schools in Pulau Pinang state, Malaysia. Each participant was required to hold a bachelor's degree in English Language Teaching (ELT) and possess no less than three years in ESL teaching. Furthermore, the teachers selected completed training and workshop sessions on CBA, whether through state or district education departmental initiatives or the professional learning community (PLC) at the school level. The demographic profiles of the participants are summarised in Table 1 below.

Table 1*Participants Demographic Profile*

Participants	Gender	Age	Educational Background	Major	Teaching experience
Teacher 1	Male	37	B.A. Linguistics	English	8
Teacher 2	Female	26	M.Ed TESL	Linguistic	4
Teacher 3	Female	31	B.Ed TESOL	English	6
Teacher 4	Female	25	B.Ed TESL	English	4
Teacher 5	Female	42	PhD Linguistics	Linguistics	16
Teacher 6	Female	45	B.Ed TESL	English	17

3.3 Data Collection Methods

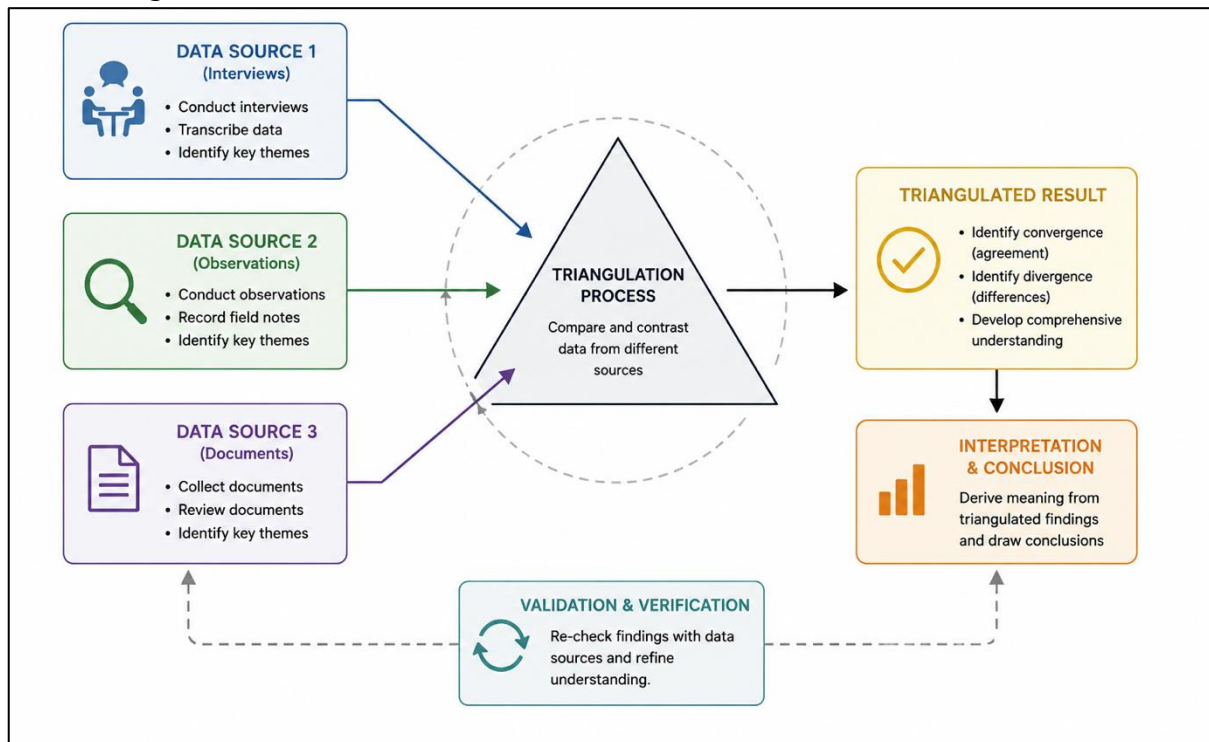
The present study employed semi-structured interviews as the primary data collection method. This approach was selected because it allows participants to articulate their experiences and perspectives in their own words, thereby generating rich and in-depth insights into the implementation of CBA grading practices (Creswell, 2002; Rahman, 2014). Semi-structured interviews are particularly suitable for exploring policy enactment at the classroom level, as they provide flexibility while maintaining a clear focus on the research objectives. An interview guide was used to ensure consistency across participants and to maintain alignment with the predetermined focus of the study (Patton, 2002). The guide enhanced the reliability and credibility of the data by incorporating pre-prepared questions and allowing relevant follow-up probes when necessary (Gesch-Karamanlidis, 2015; Lindlof & Taylor, 1994). Although the protocol was adapted from Rahman (2014), who examined curriculum reform in Malaysian primary schools, it was modified to suit the present study's focus on CBA grading practices in Malaysian ESL classrooms. The sequence of questions served as a flexible framework rather than a rigid structure, enabling the researcher to explore emerging themes while remaining within the study's scope (Merriam & Tisdell, 2015). The interview protocol used in this study contained three major parts. In Section A, biographical details concerning the teachers, such as their academic credentials and years of teaching, were captured. Section B assessed teachers' comprehension regarding applying CBA within the context of the Malaysian education system. In Section C, the researchers asked respondents to explain their grading practices, particularly the techniques and methods used during CBA. The researchers conducted each interview for approximately 20 to 30 minutes and digitally recorded the sessions for detailed analysis. They conducted the interviews in English, as the participants practiced in the ELT field.

In order to explore the alignment of teachers' grading practices with the policy mandates, this study also triangulates the data from the document analysis. In practice, document analysis was used to corroborate and interrogate interview claims by comparing teachers' reported grading routines against the specific expectations articulated in the CBA Guidebook (Ministry of Education Malaysia, 2019) and the curriculum/performance standards referenced in the DSKP/SBELC. For example, when teachers described grading through numerical accumulation and percentage conversion, such as "collecting marks" or referring to

results as “black and white,” the researchers cross-checked these practices against the guidebook’s emphasis on non-competitive, formative use of assessment information that prioritises using evidence to support follow-up actions rather than comparing students. Likewise, when participants described standard setting through panels or PLC sessions and the use of locally agreed cut scores, the document analysis examined whether such mastery decisions were explicitly anchored to curriculum-aligned performance standards as mandated in the policy guidance. In addition, school-based artefacts, including transit forms and recording templates, were analysed to identify what was actually recorded, including mastery levels, skill domains, and the frequency of recording, and whether the recorded categories reflected policy language or school-generated classifications. This process enabled the study to identify points where documents reinforced teachers’ accounts, such as expectations for continuous recording, and points where documents challenged them, particularly exam-oriented grading rationales and benchmarking practices.

In this sense, the teachers’ grading practices discovered in the semi-structured interview were compared to the CBA Guidebook (Ministry of Education Malaysia, 2019). In addition, pertinent materials used by the teachers during the grading process were also collected. A document summary form, as described by Yin (2014), was utilised for the medium of data collection. The form includes several important parts which are the date of documentation, title, description, content summary, relevance to the research field, and reflective notes by the researcher. By using this form, researchers were able to gain an overview of the document and facilitate the content analysis (Yin, 2014) to determine whether teachers’ assessment practices aligned with the policy needs. Additionally, to maintain systematic referencing, the documents collected were labelled in ascending order. For example, the first document was marked as “Document Form 1”. This labelling approach can potentially improve document identification clarity and streamline the process of linking related documents. The triangulation of these data collection methods is illustrated in Figure 4 below.

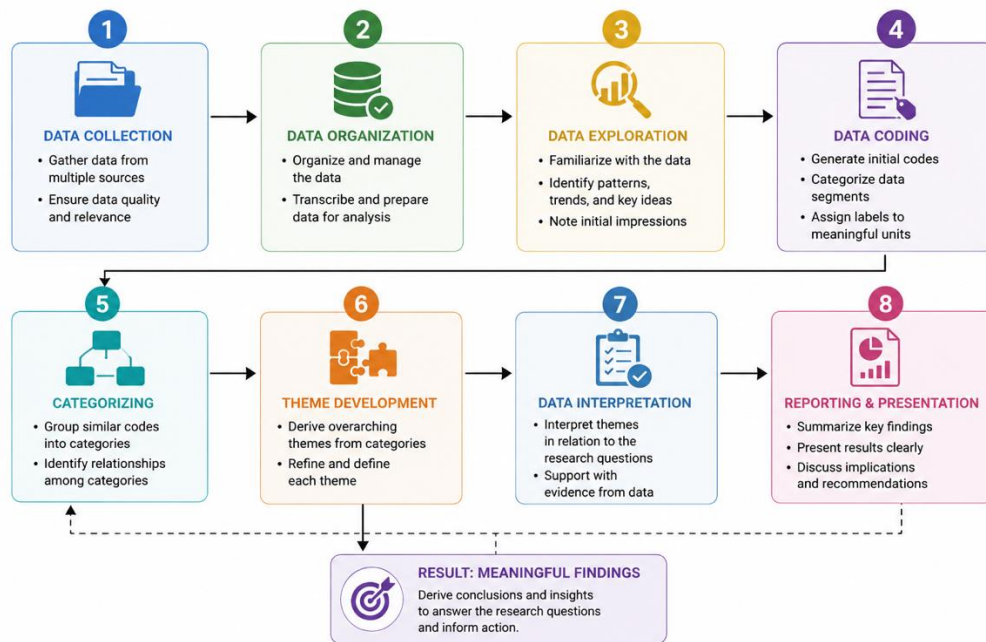
Figure 4
Data Triangulation Procedures



To gain access to the research sites, the researchers secured formal approval from the Ministry of Education's Planning and Research Development Unit (PRD) and the Pulau Pinang State Education Department (JPN). Obtaining these permissions enhanced the legitimacy of the study and enabled entry into the selected schools. Following preliminary discussions with school principals, an information letter detailing the objectives of the study and its ethical considerations was distributed. During the school visits, administrators were informed about the procedures for ensuring confidentiality and safeguarding participant data. Informed consent was subsequently obtained from all participants, with assurances of anonymity and voluntary involvement. ESL teachers, identified through recommendations from school administrators, were briefed on the study and agreed to participate. All interviews were carried out during official school hours according to a pre-arranged schedule. This organised procedure facilitated an in-depth exploration of teachers' experiences while ensuring adherence to ethical principles and methodological rigour.

3.4 Data Analysis

This study employed thematic analysis to examine teachers' grading practices in the implementation of CBA. Interview data were transcribed verbatim and analysed systematically to identify patterns and meanings embedded in participants' accounts. Figure 5 presents a visual summary of the qualitative data analysis process adopted in this study.

Figure 5*The qualitative process of data analysis*

The analysis followed the six iterative phases of thematic analysis proposed by Braun and Clarke (2006). These phases guided the analytical procedures as outlined below.

- Familiarisation with the data: An in-depth analysis was conducted to understand the insights derived from the semi-structured interviews that were carried out in the field. This procedure entails transcribing and rereading the collected data to gain an understanding of the data collected using the Happyscribe software.
- Generating initial codes: Meaningful segments were identified as data segments and assigned descriptive codes. These codes include the phrases and sentences that best convey the content's substance.
- Searching for themes: Similar codes were compiled to investigate the emergent patterns and relationships. The emerging patterns are prospective themes that may reflect the data's underlying significance.
- Reviewing Themes: Themes were reviewed and refined from the preceding sections by considering whether they make sense in relation to the entire data set. This phase involves combining, dividing, and redefining themes as necessary.
- Defining and Naming Themes: Each established theme was given a name that conveys its essence in a clear and concise manner. In addition, a comprehensive explanation was developed and provided to explain what each theme represents.
- Presenting and Discussing Results: After identifying all codes and themes, the data was systematically presented and discussed. The findings section of this study therefore discusses and presents the codes and themes that emerged from the procedures of thematic analysis.

The coding process adopted a hybrid analytical approach combining inductive and deductive logics. Initial codes were generated inductively from participants' interview transcripts to capture grading practices in teachers' own terms. Subsequently, a deductive lens was applied during theme refinement by using policy expectations as an analytical benchmark, particularly the CBA Guidebook's emphasis on formative intent, evidence-based mastery judgment, and non-comparative reporting. In this way, themes remained grounded in empirical data while being interpreted in relation to documented policy requirements. To facilitate data management and systematic coding, ATLAS.ti qualitative data analysis software was employed. The software supported data organisation, storage, retrieval, and coding processes, thereby enhancing analytical rigour and transparency (Creswell & Poth, 2023). Each participant's transcript was uploaded as a separate file to enable systematic comparison and theme development.

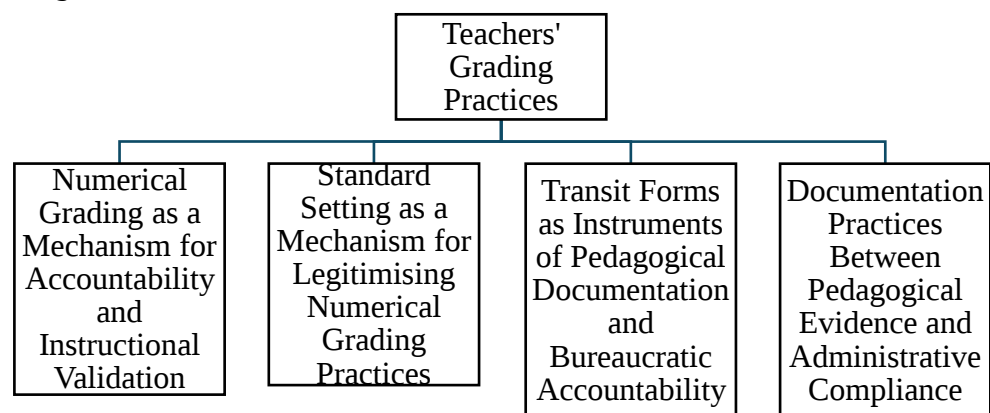
4. Findings and Discussion

4.1 Teachers' Grading Practices

A thematic analysis based on data collected in this study discovered several grading practices that the participants employed. Figure 6 below presents the grading practices discovered in this study.

Figure 6

Teachers' Grading Practices



4.1.1 Numerical Grading as a Mechanism for Accountability and Instructional Validation

In this study, 4 out of 6 teachers said they used numerical data to evaluate how well students are learning through the formative assessment of CBA. Teachers believed that the data they collected could help them determine how students were making progress in their learning while they were learning. For example, one teacher (T3) said she collected students' marks by doing a number of tasks, and then using those marks, she was able to determine if students made progress from the teaching sessions that she had with them.

For me, I gather students' scores from the quizzes administered after each lesson. I find this to be a more straightforward way of assessing their learning progress. (T3)

Another teacher (T2) noted the same thing. She indicated that this method gave specific feedback about how well the student understood the different types of materials, the topics that were taught, and the learning outcomes that the students met. It also gave quantifiable feedback to show how well the student did in particular areas, as evidenced by the use of percentages. To do this, she converted students' marks from their different tasks to percentages.

When looking at student performance from a numerical perspective can be eye-opening. We can present results in black or white; however, from my perspective as a teacher of English, I take the original marks students received for each task assigned and convert them into a percentage. Doing so allows me to see where they are currently ranked with respect to their language ability across all four skill areas... (T2)

Also, she stated that the primary aim of her percentage-based grading was to motivate her students. She found that high percentage grades can provide targets for her students to work towards, and this will help to engage her students more effectively in their learning. This caused her to believe that collecting numerical grades from students validates that her classroom instruction is being delivered appropriately.

As a result of using percentages, my student motivation levels have increased. Utilizing percentage grading provides an environment in which students' success becomes a stimulus for competition.” (T2)

According to these results, the numerical grading practices of teachers were influenced by the deeply ingrained assessment cultural beliefs that link accountability and validation of the instructional process to numerical results. Thus, it seems teachers saw numerical data as evidence of their successful instructional practice and, hence, student scores functioned not only as indicators of student learning, but also as validation of teacher effectiveness. This twofold use of grades continues to perpetuate traditional, high-stakes assessment logics, while the policy framework supports formative, process-oriented assessment practices. When compared to the recommendations of the CBA Guidebook (Ministry of Education Malaysia, 2019), these grading practices are not consistent with what policymakers require. As a result, these teachers' grading practices are not in alignment with the CBA Guidebook's recommendation to maintain a non-threatening environment, provide no external motivation by comparing student performance, and communicate student accomplishments to guide the needed changes to improve the educational process.

Teachers should create a relaxed, enjoyable, and supportive classroom environment, ensuring that interactions between teachers and students, among students, and with learning materials occur in a meaningful and engaging manner.

(Ministry of Education Malaysia, 2019, p. 39)

Asikainen et al. (2013) further supported the notion that focusing solely on grades in classroom assessments does not necessarily reflect the true learning outcomes and students' abilities. They argued that an overemphasis on numerical grades can lead to a narrow view of student achievement, where the depth of understanding, critical thinking skills, and practical application of knowledge may be overlooked. Their research suggests that grades alone are insufficient in capturing the full spectrum of a student's educational experience and intellectual development. Rather, they propose to incorporate feedback, self and peer assessments as part of the evaluation process for the learner's progress tracking, in addition to more formal methods of evaluation, to offer a complete picture of a student's competencies and learning development.

4.1.2 Standard Setting as a Mechanism for Legitimising Numerical Grading Practices

This study also confirmed the existence of standard setting procedures in relation to grading from the evidence of the significant reliance on numeric accumulation by teachers when grading students. A standard setting procedure is a way in which teachers create and agree on a common understanding of how their students perform (Sondergeld et al., 2018). This process is done by assessing and comparing students' work against predetermined formal standards and predetermined success indicators at different levels of performance. Standard setting is a practice which can encompass an eclectic range of approaches and methods; for example, teachers use their own decision-making processes regarding cut scores for alignment with proficiency scales (Kampa et al., 2019). Wiggins and McTighe (2005) describe standard setting as a way to establish a framework of support for teachers' professional judgement of their students' level of mastery of content. Concerning this research, three of the six teachers whom being interviewed indicated that they performed standard-setting processes, often to create a boundary for grading students or to confirm their grades were within the boundaries of standardization. Teacher 1's testimony illustrated this point as he stated that standard-setting is used to create fairness when evaluating students' mastery of content across teachers in the same school.

We conducted standard-setting discussions within the panel by presenting our assessment data and comparing it with those of other teachers. This process was carried out to ensure a more balanced distribution of scores across teachers. (T1)

Teacher 4 used the standard-setting method for evaluating students through her school's Professional Learning Community (PLC) group. While PLCs typically refer to collaborative structures within schools where teachers engage in ongoing learning and collaboration (Chen, 2020), PLCs also create a supportive environment in which teachers can collaboratively reflect on their practices, share resources, and discuss their strategies for improvement. Teachers can improve their knowledge, refine their instructional strategies, and intentionally modify their practices to impact student learning by implementing reflective practice during their professional development. According to Penner-Williams et al. (2017), reflective practice will enhance the effectiveness of instructional strategies and facilitate professional development and lifelong learning. Teacher 4 indicated that her school's PLC determined the cut scores that

were used to rate students' assessments in the classroom and assigned the respective performance categories associated with those cut scores.

...the teachers would meet to determine the standards for each group based on the students' grades. Following the discussion, I recorded my students' mastery levels according to the standards collectively agreed upon by all teachers ... (T4)

This study shows that the establishment of academic standards is a subjective process but can also be controversial when it comes to classroom assessments. Different schools may have varying cut score-setting practices which, in turn, affect their assessment practices. Some teachers in this study continued to create their cut scores without using the performance standards of SBELC. The results indicate there is a gap between expectations of policy and classroom practices as established in the CBA Guidebook (Ministry of Education Malaysia, 2019), which states that decisions about mastery at the claim level must reflect the curriculum standards. An illustration of this is presented in the next excerpt.

The procedures of determining students' mastery level should depend on the teacher's autonomy and should be carried out in accordance with the requirements and specifications of the subject's Curriculum and Assessment Standard Document (DSKP).

(Ministry of Education Malaysia, 2019, p. 39)

The teachers' dependence on mark accumulation and numerical grades in the execution of CBA showcases a marked deviation from the policy's intent. Focusing on recording students' marks reflects an approach more aligned with the previously dominant exam-centred evaluation framework, where a student's relative standing was the primary metric informing their value, subsequently resulting in score benchmarking. Examinations and such competitions are not in accordance with what was advocated in the CBA Guidebook, which highlights...

Assessment information should not be used for comparing or competing among students. Instead, it should serve as a tool to help schools and parents plan appropriate follow-up actions to support and improve student learning mastery and achievement.

(Ministry of Education Malaysia, 2019, p. 5)

The variability of cut scores demonstrates the amount of subjectivity involved in standard setting in a classroom context (Townsley & Varga, 2018); therefore, if grading procedures differ among schools, both fairness and consistency are called into question. Furthermore, the fact that there is no consistent reference to national performance standards suggests that "standardisation" is limited to the location (school) but not to an overall system-wide level. Thus, there is a possibility of fragmented implementation of student performance levels. Overall, this study indicates that teachers are involved with standard-setting activities, but the standard-setting processes are primarily based on scores and are sensitive to the score

distributions. Standard-setting provides teachers with a way to incorporate criterion-referenced principles of CBA policy. However, standard-setting seems to be functioning primarily as a way for teachers to legitimise the numerical grading process in their classrooms. Overall, CBA is being implemented through the lens of traditional assessment culture through the use of quantification, benchmarking, and perceived objectivity

4.1.3 Documentation Practices Between Pedagogical Evidence and Administrative Compliance

Based on the outcomes of the thematic analysis, two teachers asserted that a fundamental approach to evaluating students' mastery levels involves the systematic collection of evidence derived from students' academic endeavours. Teachers expressed the belief that the compilation of students' work constitutes an indispensable step in effectively assessing and grading their overall performance. This process, they argued, allows for a comprehensive monitoring of students' progress throughout the learning journey. This sentiment was vividly articulated by Teacher 5, who described her diligent practice of collecting students' work during classroom assessments, highlighting the significance of this method in shaping informed decisions about students' mastery skills. Teacher 5 elaborated on her approach by stating that,

I always collect my students' work. For me, it's easier to have this kind of evidence kept closely in the file so that I can refer back to monitor my students' performance before deciding their mastery skills. (T5)

This reflective practice underscores the value of maintaining a tangible record of students' work as a reference point for gauging and understanding their academic achievements. In a similar vein, Teacher 6 adopted a comparable strategy in assessing students' mastery levels. She emphasized the ease and efficiency brought about by the systematic collection and documentation of students' work. According to Teacher 6, this practice not only streamlines the grading process but also facilitates the establishment of performance standards tailored to the diverse capabilities of students in the classroom. Teacher 6 stated,

Through documenting my students' work, I believe it helps me a lot in terms of knowing how each student differs among them. This can help me to see how to set the performance standard among students in my classroom. (T6)

This collective insight from teachers aligns with the findings from Zamri and Hamzah (2019), who conducted a study on primary teachers' competency in implementing classroom assessment. Their research emphasized the crucial role of collecting evidence about student learning during classroom assessments in providing an overview of students' performance. This, in turn, enhances teachers' competency in making professional judgments about students' mastery levels. The convergence of these perspectives underscores the significance of evidence-based assessment practices in fostering effective teaching and learning environments.

...teachers should collect an adequate amount of students' work and conduct observations to gain a comprehensive picture to assist in a making an assessment of a student...

(Ministry of Education Malaysia, 2019, p. 38)

Despite participants affirming their engagement in the collection of student work and evidence for performance assessment, this study revealed an alternate perspective among minority teachers. These educators asserted that such practices were undertaken solely to comply with the expectations of school administrators. Teachers expressed a sense of reluctance, indicating that they engaged in these activities primarily in response to demands from higher-ranking stakeholders who sought to monitor teachers' work conduct. Some teachers voiced the opinion that the collection of students' work and evidence did not significantly contribute to their professional judgment.

"... I was asked to do this years ago, at the time I didn't think it would help me very much at all in assessing my students and in effect keeping a record of what we did in class... (T6)

"...when I collect evidence from the student's work it doesn't really help me assess how well they are performing, but I find that doing in-class assessments helps me assess students in real time, and that is what is important to me... (T5)

This contradiction suggests that documentation may function in two competing ways: 1) as a pedagogical tool for informed judgment, and as an accountability mechanism for administrative surveillance. Such duality reflects what policy enactment literature describes as compliance-driven implementation, where teachers perform required procedures without fully internalising their pedagogical rationale. The perception that documentation primarily serves monitoring purposes indicates a possible misalignment between policy intention and classroom reality. Importantly, teachers' preference for "real-time performance" highlights a deeper epistemological issue: what counts as valid evidence of learning? For some participants, spontaneous classroom interaction appeared more authentic than archived written artefacts. This suggests that while teachers may comply with documentation requirements, their underlying beliefs about assessment validity remain anchored in immediate classroom observation rather than portfolio-based evidence accumulation. Therefore, the divergence in perspectives is not merely attitudinal—it signals varying levels of assessment literacy and differing interpretations of professional accountability. The findings imply that documentation practices in CEFR-aligned CBA are shaped not only by policy mandates but also by teachers' beliefs about evidence, authenticity, and administrative oversight.

4.1.4 Transit Forms as Instruments for Pedagogical Documentation and Bureaucratic Accountability

This research also noted that three out of six teachers employed a rubric known as the 'transit form' to aid the marking processes within the CBA structure. This form is designed to capture assessment data over a given length of time, cut to fit scaffolded goals; thus, formative evaluations are structured in a manner where they can be aligned with the objectives sought.

Generally, the form has a table of the learning outcomes and columns for each student's mastery level of these outcomes throughout the year. For example, Teacher 2 mentioned that she received a transit form with two key columns for yearly assessment data. Because CBAs were documented and reported biannually at her school, this facilitated the teachers to benchmark and reflect on students' progress during the assessment cycle.

We have a document called the transit form, which lists the names of the students and gives a column for recording their level of mastery over time. The administrator required us to record the students' performance both at the beginning of each year and at the end of each year so that they can track the students' progress... (T2)

Additionally, Teacher 3 received a recording template to capture her students' performance throughout the assessment process. Where she works, the form mentioned above (see Figure 7 below) has the status of an official reference document for the school administrators. As the teachers did not have access to the standardised document, it became a sustaining personal record system to measure and monitor the progress of students over the years.

Figure 7
Sample of Transit Form used by Teachers

BORANG TRANSIT PENTAKSIRAN MATA PELAJARAN BAHASA INGGERIS TAHUN 1									
NAMA SEKOLAH: _____					TEMIN				
ALAMAT SEKOLAH: _____					INTUT				
KOD SEKOLAH: _____					2050				
TARIKH PELAPORAN: _____									
NAMA GURU MATA PELAJARAN: _____					KELAS: _____				
TEMPOH TAKSIRAN: _____									
BIL	NAMA MURID	STANDARD PEMBELAJARAN							
		LISTENING							
		1.1.1	1.2.1	1.2.2	1.2.3	1.2.4			
	TARIKH PENTAKSIRAN								
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									

Based on her statement, the school's transit form was organized according to the skills delineated in the curriculum's scope and sequence chart, including the learning outcomes and benchmarks. Such alignment was meant to guarantee that teachers evaluated the appropriate knowledge and skills presumed to have been acquired by the students. Therefore, teachers were obliged to carry out ongoing formative assessments of students by evaluating their performance against the criteria set out in the SBELC content and learning standards.

The school administration has given me a template for keeping track of my students' progress. This template is divided into categories according to the specific skills I want to assess my students on. It helps me keep accurate records so I can monitor how well my students are doing over time. (T3)

This practice aligns with the SBELC recommendation that CBA recording should be done continuously to provide teachers with an overview of students' performance from the beginning to the end of the year.

The recording activity is carried out by the teacher from time to time to ensure that comprehensive information about students' mastery of learning is collected and easily referenced. The information gathered allows the teacher to make professional judgments to determine the students' level of mastery within a group of Content Standards and Learning Standards.

(Ministry of Education Malaysia, 2019, p. 46)

Referring to the previously described practices, the implementation of CBA has sparked a heated debate among teachers. Teachers may become fixated on administrative processes and documentation at the expense of valuable assessment strategies that would otherwise enhance student learning. The transit form and accompanying template were seen by many as aids to assist lesson planning but also added to the workload. This was highlighted by the comment of Teacher 3, who reported that they were not happy with the CBA process, calling it a mindless exercise, especially given the disproportionate amount of work required to collect and administratively clean the data for the assessments.

...it requires extensive and laborious effort you know, particularly during the data entry and preparation stages. These demands are set forth by higher authorities... (T3)

Moreover, she conveyed that the focus placed on capturing assessment data has distracted her from constructing and improving effective assessment strategies for her classroom. This issue is deeply connected with the type of information that the teacher uses to verify the accuracy of her decisions regarding the evaluation, thus creating a barrier to the use and management of assessment information systems.

They consistently critique our data, highlighting parts deemed unreliable while approving others. Consequently, there are occasions where I must alter the data to align with their requirements... (T3)

This statement shifts the function of the transit form from a formative assessment tool to an instrument of performative accountability. Rather than facilitating authentic reflection on student learning, the documentation process appears to prioritise data conformity. The need to “alter the data” suggests the presence of external pressures that may compromise professional autonomy and the integrity of assessment evidence. Thus, the transit form embodies a dual function: 1) as the pedagogical instrument which helps in structuring evidence collection and

aligning assessment with learning standards; 2) a bureaucratic mechanism which helps in enforcing compliance, standardisation, and administrative surveillance. This duality reflects broader tensions in CBA reforms, where documentation intended to enhance assessment literacy may instead intensify workload and shift focus toward data management rather than instructional improvement. As Cochran-Smith (2003) cautions, an overemphasis on grading data risks narrowing teachers' engagement with the richer dimensions of student learning. Importantly, the teachers' dissatisfaction does not indicate rejection of assessment principles *per se*; rather, it signals friction in policy enactment. The issue lies not in recording evidence, but in how documentation practices are operationalised, monitored, and interpreted within institutional hierarchies. When recording becomes a compliance ritual rather than a reflective practice, its formative potential is diminished.

5. Conclusion

This study explored ESL teachers' grading practices in CBA and examined the degree to which these practices aligned with policy mandates. The findings show that teachers employed multiple strategies to document and judge student learning, including the use of recording templates, documentation of student evidence, and collaborative standard-setting practices. However, a consistent pattern of misalignment emerged where grading practices frequently leaned toward exam-oriented norms, particularly through numerical mark accumulation and benchmarking, rather than criterion-referenced mastery decisions anchored in curriculum standards. These misalignments are significant because they risk diluting the formative intent of CBA by shifting attention from feedback and learning progression to score-focused reporting and competition. In addition, inconsistent standard-setting practices and locally determined cut scores may contribute to variation in mastery judgments across teachers and schools, potentially weakening fairness and the credibility of reported achievement for parents and stakeholders.

The findings of this study generate important implications for consolidating the coherence and long-term sustainability of CBA reform within the Malaysian education system. Although the Ministry of Education Malaysia promotes a formative and standards-based approach through the CBA framework, teachers' continued reliance on numerical marks, percentage conversion, and locally negotiated cut scores suggests that classroom implementation remains influenced by examination-oriented grading practices. This suggests that policy articulation alone is insufficient to recalibrate entrenched examination-oriented grading logics. To strengthen alignment, policy instruments should incorporate greater operational specificity, including annotated exemplars of mastery-level determinations anchored explicitly to SBELC performance descriptors, structured moderation protocols that support cross-teacher calibration, and decision-making heuristics that clarify how professional judgement can be exercised without defaulting to percentage thresholds. Simultaneously, professional development must move beyond procedural compliance and documentation training toward sustained, practice-embedded assessment literacy development, particularly in cultivating teachers' capacity to interpret diverse forms of student evidence longitudinally and to render defensible, standards-referenced mastery judgements. At the institutional level, school leadership and monitoring systems should be reoriented from audit-driven documentation compliance toward evaluating the quality of feedback cycles, instructional adaptation, and intervention planning, thereby reinforcing the formative integrity of assessment

processes. Without such multi-level recalibration, policy clarity, professional capacity building, and leadership mediation, there is a risk that structural reform away from high-stakes examinations will reproduce examination epistemologies under a different administrative architecture, limiting the transformative potential of CBA.

Several limitations delimit the interpretive scope and transferability of this study. The qualitative case study design, involving six ESL teachers from three public primary schools within a single Malaysian state, enabled contextual depth and analytic coherence but necessarily constrains statistical generalisability; thus, the findings should be understood as analytically transferable rather than nationally representative accounts of grading enactment under CBA. Furthermore, the study was confined to public primary schools operating under uniform mandates issued by the Ministry of Education Malaysia, and therefore does not capture potential variations across primary, private, or international school contexts where differing accountability regimes, leadership configurations, and stakeholder expectations may shape assessment practices differently. Although document analysis was used for triangulation, the primary data source consisted of teachers' self-reported accounts, which may reflect policy-aligned rationalisations rather than fully enacted, real-time grading decisions; consequently, micro-level classroom interactions and spontaneous formative judgements were not directly observed. In addition, given that CBA reform remains relatively recent within broader systemic reform cycles, current practices may represent transitional enactments rather than stabilised professional norms. In light of these limitations, future research should pursue comparative multi-site investigations across diverse school types and regions to examine contextual mediation of grading alignment, adopt mixed methods designs integrating assessment literacy measures with observational data to enhance explanatory robustness, and conduct longitudinal studies tracking teachers over time to determine whether professional judgement becomes more consistently anchored to standards as reform matures. Further inquiry into students' perceptions of mastery-based grading would also enrich understanding of how shifts in teacher judgement influence learner motivation, self-regulation, and classroom climate, thereby advancing theoretical conceptualisation of assessment reform as a dynamic, multi-level process rather than a discrete policy event.

6. Declaration of AI Use

The authors declare that ChatGPT 5.0 was used in generating the data triangulation procedures diagrams and the qualitative process of data analysis (Figures 4 and 5) to show the steps and procedures on the methodology in the present study only. No other parts of the manuscript were generated by AI tools.

7. References

- Abdul Razak, R., Mat Yusoff, S., Leng, C. H., & Mohamad 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. <https://doi.org/10.1371/journal.pone.0293325>
- Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 20–34. <https://doi.org/10.1080/0969594X.2012.736364>

- Arumugham, K. S. (2020). Kurikulum, pengajaran dan pentaksiran dari perspektif pelaksanaan pentaksiran bilik darjah. *Asian People Journal (APJ)*, 3(1), 152–161. <https://doi.org/10.37231/apj.2020.3.1.175>
- Asikainen, H., Parpala, A., Virtanen, V., & Lindblom-Ylänne, S. (2013). The relationship between student learning process, study success and the nature of assessment: A qualitative study. *Studies in Educational Evaluation*, 39(4), 211–217. <https://doi.org/10.1016/j.stueduc.2013.10.008>
- Black, P. (2015). Formative assessment—An optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, 22(1), 161–177. <https://doi.org/10.1080/0969594X.2014.999643>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1). <https://doi.org/10.1080/0969595980050102>
- Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125–230. <https://doi.org/10.3102/00346543073002125>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Brookhart, S. M., & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education*, 16(1), 27–54. https://doi.org/10.1207/S15324818AME1601_2
- Chen, L. (2020). A historical review of professional learning communities in China (1949–2019): some implications for collaborative teacher professional development. *Asia Pacific Journal of Education*, 40(3), 373–385.
- Chin, H., Thien, L. M., & Chiew, C. M. (2019). The reforms of national assessment in Malaysian education system. *Journal of Nusantara Studies (JONUS)*, 4(1), 93–111. <https://doi.org/10.24200/jonus.vol4iss1pp93-111>
- Cochran-Smith, M. (2003). Assessing assessment in teacher education. *Journal of Teacher Education*, 54(3), 187–191. <https://doi.org/10.1177/0022487103054003001>
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.
- Creswell, J. W., & Poth, C. N. (2023). *Qualitative Inquiry and Research Design* (5th ed.). SAGE Publications.
- Eli, A., & Yamat, H. (2021). A preliminary study on classroom-based assessment for English speaking performance among Year 5 pupils in an urban school. *International Journal of Advanced Research in Education and Society*, 3(4), 45–51.
- Gesch-Karamanlidis, E. (2015). Reflecting on Novice Qualitative Interviewer Mistakes. *The Qualitative Report*, 20(5). <https://doi.org/10.46743/2160-3715/2015.2145>
- Ghavifekr, S., Kunjappan, T., Ramasamy, L., & Anthony, A. (2016). Teaching and learning with ICT tools: issues and challenges from teachers' perceptions. *MOJET Online Journal of Educational Technology*, 4(2), 38–57.
- Hanapi, M. H. M., Vadivelu, S. D., Zakaria, N., & Othman, M. S. (2022). Primary school classroom assessment (PBD) implementation evaluation in Shah Alam, Selangor. *KQT EJournal*, 2(2), 1–9. <http://ejurnal.kqt.edu.my/index.php/kqt-ojs/article/view/57>

- Hay, A. (2024). What may be: policy enactment in education, a new conceptual framework with actor-network theory. *Journal of Education Policy*, 1–20. <https://doi.org/10.1080/02680939.2024.2411989>
- Kampa, N., Wagner, H., & Köller, O. (2019). The standard setting process: Validating interpretations of stakeholders. *Large-Scale Assessments in Education*, 7(1). <https://doi.org/10.1186/s40536-019-0071-8>
- Khory, H. H. M., Rahman, M. N. A., & Zailaini, M. A. (2021). Pengurusan pentaksiran bilik darjah mata pelajaran Bahasa Arab berasaskan keperluan pembelajaran murid. *JuPiDi: Jurnal Kepimpinan Pendidikan*, 8(2), 41–57.
- Lindlof, T. R., & Taylor, B. C. (1994). *Qualitative communication research methods*. SAGE Publications.
- Machin, S., McNally, S., & Ruiz-Valenzuela, J. (2020). Entry through the narrow door: The costs of just failing high stakes exams. *Journal of Public Economics*, 190, 104224. <https://doi.org/10.1016/j.jpubeco.2020.104224>
- Marnizam, F., & Ali, S. (2021). Evaluation of the implementation of classroom assessment (PBD) among primary school mathematics teachers. *Jurnal Pendidikan Sains Dan Matematik Malaysia*, 11(2), 81–94. <https://doi.org/10.37134/jpsmm.vol11.2.7.2021>
- Mat Yusoff, S., Leng, C. H., Abdul Razak, R., Basal, M. H., & Mohamad Marzaini, A. F. (2024). Conceptions of assessment: Perceptions of physical and health education teachers in Selangor, Malaysia. *Pertanika Journal of Social Sciences & Humanities*, 32(1), 189–216. <https://doi.org/10.47836/pjssh.32.1.09>
- Mat Yusoff, S., Mohamad Marzaini, A. F., 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. <https://doi.org/10.37134/mjm.vol12.2.6.2023>
- Mat Yusoff, S., Mohamad Marzaini, A. F., Jusoh, A. M. M., Zynuddin, S. N., Basal, M. H., & Zakaria, N. (2025a). Investigating the interrelationship between teachers' grading practices and classroom-based assessment strategies among Malaysian preservice teachers. *Asian Journal of University Education (AJUE)*, 21(1), 177–199. <https://doi.org/10.24191/ajue.v21i1.5622>
- Mat Yusoff, S., Mohamad Marzaini, A. F., Zakaria, N., & Lijie, H. (2025b). 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. <https://doi.org/10.22452/mojem.vol13no1.5>
- McMillan, J. H., & Workman, D. J. (1998). *Classroom assessment and grading practices: A review of the literature* (ERIC Document No. ED453263). ERIC. <https://eric.ed.gov/?id=ED453263>
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.
- Ministry of Education Malaysia. (2019). *Panduan pentaksiran bilik darjah* [Classroom-based assessment guidebook] (2nd ed.). Bahagian Pembangunan Kurikulum, Kementerian Pendidikan Malaysia.
- Mohamad Marzaini, A. F., Haslee Sharil, W. N. E., Supramaniam, K., & Mat Yusof, S. (2023). Unravelling the factors impinging the implementation of classroom-based assessment in Malaysian ESL classrooms: Voice from practitioners. *International Journal of*

- Advanced Research in Education and Society*, 5(3), 370–393.
<https://doi.org/10.55057/ijares.2023.5.3.35>
- Mohamad Marzaini, A. F., & Mat Yusoff, S. (2026). A qualitative examination of misalignments in ESL teachers' grading practices within classroom based assessment in Malaysian schools. *Discover Education*, 5(51). <https://doi.org/10.1007/s44217-025-01009-0>
- Mohamad Marzaini, A. F., Yusof, S. M., Zakaria, N., Basal, M. H., Haslee Sharil, W. N. E., & Supramaniam, K. (2024). A discovery on Malaysian ESL teachers' adherence to policy mandates in classroom-based assessment practices. *Malaysian Journal of Learning & Instruction*, 21(1), 249–284. <https://doi.org/10.32890/mjli2024.21.1.9>
- Mohd Faudzi, N. H., & Ahmad, A. (2022). *Kesediaan guru sejarah melaksanakan pentaksiran bilik darjah dalam pengajaran dan pembelajaran*. <https://www.icbe.my/wp-content/uploads/2022/04/Manuscript-ICBE-Nurul-Huda-1-2.pdf>
- Mohd Isa, A., Mydin, A.-A., & Abdullah, A. G. K. (2022). Autonomy in classroom assessment: Teachers' perceptions on the abolition of Level 1 biennial examinations in Malaysian national schools. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 7(3), e001331. <https://doi.org/10.47405/mjssh.v7i3.1331>
- Patallion Jr, A. P. (2016). The implications of learning theories to assessment and instructional scaffolding techniques. *American Journal of Educational Research*, 4(9), 695–700. <https://doi.org/10.12691/education-4-9-9>
- Patton, M. Q. (2002). *Qualitative research & evaluation methods*. SAGE.
- Penner-Williams, J., Díaz, E. I., & Gonzales Worthen, D. (2017). PLCs: Key PD component in learning transfer for teachers of English learners. *Teaching and Teacher Education*, 65, 215–229. <https://doi.org/10.1016/j.tate.2017.03.014>
- Primus, V. W., & Mosin, M. (2021). Pengurusan pentaksiran bilik darjah murid berkeperluan khas dalam program pendidikan inklusif di sekolah rendah. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*, 6(8), 59–68. <https://doi.org/10.47405/mjssh.v6i8.903>
- Rahman, N. H. (2014). *From curriculum reform to classroom practice: An evaluation of the English primary curriculum in Malaysia* (Doctoral dissertation, University of York).
- Sondergeld, T. A., Stone, G. E., & Kruse, L. M. (2018). Objective standard setting in educational assessment and decision making. *Educational Policy*, 34(5), 735–759. <https://doi.org/10.1177/0895904818802115>
- Sun, Y., & Cheng, L. (2014). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343. <https://doi.org/10.1080/0969594X.2013.768207>
- Tee, D. D., & Ahmed, P. K. (2014). 360 degree feedback: An integrative framework for learning and assessment. *Teaching in Higher Education*, 19(6), 579–591. <https://doi.org/10.1080/13562517.2014.901961>
- Thiede, K. W., Brendefur, J. L., Carney, M. B., Champion, J., Turner, L., Stewart, R., & Osguthorpe, R. D. (2018). Improving the accuracy of teachers' judgments of student learning. *Teaching and Teacher Education*, 76, 106–115. <https://doi.org/10.1016/j.tate.2018.08.004>
- Thompson, J. L. (2018). *Adult education for a change*. Routledge.

- Townsley, M., & Varga, M. (2018). Getting high school students ready for college: A quantitative study of standards-based grading practices. *Journal of Research in Education*, 28, 92–112.
- Widiastuti, I. A. M. S. (2018). Teachers' classroom assessment and grading practices. *SHS Web of Conferences*, 42, 00052. <https://doi.org/10.1051/shsconf/20184200052>
- Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. ASCD.
- Yin, R. K. (2014). *Case study research*. SAGE Publications.
- Yusoff, S. M., Lijie, H., & Basal, M. H. (2025). Investigating the influence of pedagogical content knowledge on formative assessment practices in Islamic Religious Education of Malaysian secondary schools. *International Journal of Asia-Pacific Studies*, 21(2), 89–113. <https://doi.org/10.21315/ijaps2025.21.2.5>
- Zamri, N. M., & Hamzah, M. I. B. (2019). Teachers' competency in implementation of classroom assessment in learning. *Creative Education*, 10(12), 2939–2946. <https://doi.org/10.4236/ce.2019.1012218>